Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет __*Інфокомунікацій*_____
(повна назва)
Кафедра _____*Інформаційно-мережної інженерії*_____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____другий (магістерський)_____

Застосування методів контурного аналізу для підвищення продуктивності технологій кодування відеоінформації
(тема)

Виконав:
студент _2_ курсу, групи __*ІМІм-20-2*___
Галушка С.В.
(прізвище, ініціали)

Спеціальність _*172. Телекомунікації та радіотехніка*_____
(код і повна назва спеціальності)
Тип програми *освітньо-професійна*
(освітньо-професійна або освітньо-наукова)
Освітня програма _*Інформаційно-мережна інженерія*_____
(повна назва освітньої програми)

Керівник ____*доц.Харченко Н.А.*_____
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ *Безрук В.М.*
(підпис) (прізвище, ініціали)
2021 р.

Не містить відомостей, заборонених
до відкритого публікування

Керівник_____ /*Н.А.Харченко*

Студент _____ / *С.В. Галушка*

Харківський національний університет радіоелектроніки

Факультет *Інфокомунікацій*

Кафедра *Інформаційно-мережної інженерії*

Рівень вищої освіти другий (магістерський)

Спеціальність *172 Телекомунікації та радіотехніка*
(код і повна назва)

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

Освітня програма *Інформаційно-мережна інженерія*
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«10» *листопада* 2021 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові *Галущі Сергію Вячеславовичу*
(прізвище, ім'я, по батькові)

1. Тема роботи *Застосування методів контурного аналізу для підвищення продуктивності технологій кодування відеоінформації*

затверджена наказом університету від 8 листопада 2021 р. № 1674 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 16 грудня 2021 р.

3. Вихідні дані до роботи *Розглянути та проаналізувати існуючі методи кодування відео. Виявити властиві їм недоліки. Обгрунтувати необхідність урахування рівня семантичної значимості оброблюваних фрагментів відеокадрів. Дослідити підходи до виявлення контурної інформації у межах окремих кадрів на базі групи порогових методів, що дозволяють виявити факт присутності контурних даних та групи методів, що базуються на використання ковзаючи масок. Виконати порівняльний аналіз даних груп методів.*

4. Перелік питань, що потрібно опрацювати в роботі *Вступ*

1. Застосування відеоінфіормації у сучасному світі та ключові підходи до її обробки на рівні кінцевих вузлів

2. Особливості обробки відеоінформаційних потоків у рамках існуючого технологічного базису

3. Розробка підходів до кодування найбільш семантично значимих ділянок відеоінформаційного потоку

4. Порівняльний аналіз методів виявлення контурів загальної та спеціалізованої груп

Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) _____
Слайди у форматі Power Point (назва та мета роботи, вимоги до відеокодеків та кодованого відеопотоку, важливість урахування семантично значимих областей кадру, метод виявлення факту присутності контуру у фрагментах методи ковзаючої маски, оцінка продуктивності методів контурного аналізу)

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	*Ознайомлення із завданням. Уточнення ТЗ.*		
2	*Підбір літератури за темою роботи.*		
3	*Виконання розділу 1*		
4	*Виконання розділу 2*		
5	*Виконання розділу 3*		
6	*Виконання розділу 4*		
7	*Оформлення презентаційного матеріалу,*		
	підготовка до захисту у ЕК		

Дата видачі завдання *10 листопада* 2021 р.

Студент _____
(підпис)

Керівник роботи _____ *доц. Харченко Н.А.*
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 78 с., 23 рис., 23 джерела, 1 додаток

MOTION JPEG, КОВЗАЮЧА МАСКА, ЛІНІЙНА ФІЛЬТРАЦІЯ, ВЕКТОР ГРАДІЄНТУ, СЕМАНТИЧНА ЗНАЧИМІСТЬ, ОПЕРАНД СОБЕЛЯ

Об'єкт дослідження – методи виявлення контурів у зображенням, спрямовані на підвищення ефективності технологій кодування відео.

Мета роботи – дослідження принципу функціонування та порівняльний аналіз методів виявлення контурів.

Здійснюється аналіз функціонування технологій кодування відеоінформації. Доводиться важливість урахування семантичної значимості оброблюваних фрагментів. Досліджуються порогові методи виявлення факту присутності контурів та методи на базі ковзаючи масок. Виконується оцінка продуктивності обох підходів до виявлення контурної інформації.

THE ABSTRACT

Explanatory note: 78p., 23 fig., 23 sources, 1 app.

MOTION JPEG, SLIDING MASK, LINEAR FILTRATION, GRADIER VECTOR, SEMANTIC SIGNIFICANCE, SABEL OPERAND

The object of research - methods of detecting contours in the image, aimed at improving the efficiency of video encoding technologies.

The purpose of the work is to study the principle of functioning and comparative analysis of contour detection methods. The analysis of functioning of technologies of coding of video information is carried out. The importance of taking into account the semantic significance of the processed fragments is proved. Threshold methods for detecting the presence of contours and methods based on sliding masks are investigated. The performance of both approaches to the detection of contour information is evaluated.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ 11
ВСТУП 12
1 ЗАСТОСУВАННЯ ВІДЕОІНФІОРМАЦІЇ У СУЧАСНОМУ СВІТІ ТА КЛЮЧОВІ ПІДХОДИ ДО ЇЇ ОБРОБКИ НА РІВНІ КІНЦЕВИХ ВУЗЛІВ..... 15
1.1 Сьогод’нішня роль відеоінформації у суспільстві 15
1.2 Огляд ряду існуючих технологій кодування відео 17
1.2.1 MPEG-1 17
1.2.2 MPEG-2 18
1.2.3 MPEG-4 та рекомендації H.26* 19
1.2.4 Сімейство M-JPEG 21
1.2.5 Фрактальні методи кодування 22
1.3 Ключове завдання технологій кодування відеоінформації 23
1.4 Недоліки існуючих відеокодеків 26
2 ОСОБЛИВОСТІ ОБРОБКИ ВІДЕОІНФОРМАЦІЙНИХ ПОТОКІВ У РАМКАХ ІСНУЮЧОГО ТЕХНОЛОГІЧНОГО БАЗИСУ 29
2.1 Ключові вимоги до процесу кодування динамічних відеоінформаціних ресурсів 29
2.2 Специфічні особливості динамічних відеоінформаційних потоків 31
2.3 Проблематика формування CBR-потоків відео 35
2.4 Умови раціонального розподілу інформаційної інтенсивності усередині кодованих відеокадрів..... 38
2.5 Висновки за розділом 42
3 РОЗРОБКА ПІДХОДІВ ДО КОДУВАННЯ НАЙБІЛЬШ СЕМАНТИЧНО ЗНАЧИМИХ ДІЛЯНОК ВІДЕОІНФОРМАЦІЙНОГО ПОТОКУ 44
3.1 Алгоритми пошуку контурних ділянок у кадрі 44
3.2 Сутність алгоритмів загальної групи виявлення контурів 45
3.2.1 Декомпозиція відеопотоку до різних рівнів представлення для виявлення найбільш ефективних для виявлення контурної інформації 46
3.2.2 Процедура виявлення контурів на рівні фрагментів кадру 47
3.3 Виявлення контурів у зображеннях на базі алгоритмів спеціалізованої групи..... 52
3.3.1 Операнд Робертса..... 56
3.3.2 Операнд Собеля..... 58
3.3.3 Операнд Превіта..... 59
4 ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ КОНТУРІВ ЗАГАЛЬНОЇ ТА СПЕЦІАЛІЗОВАНОЇ ГРУП 62

4.1 Критерії порівняння методів виявлення контурів 62
4.1.1 Розрахунок кількості необхідних арифметичних операцій для обробки фрагменту зображення 62
4.1.2 Порівняння підходів до виявлення контурної інформації з позиції достовірності одержуваних даних 68
4.1.3 Порівняння підходів до виявлення контурів за функціональними відмінностями 71
ВИСНОВКИ 74
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ 77
ДОДАТОК А СЛАЙДИ ПРЕЗЕНТАЦІЇ 79
ДОДАТОК Б ТЕЗІСИ 87

ПЕРЕЛІК СКОРОЧЕНЬ

ATSC – (Advanced Television Systems Committee) – набір стандартів для цифрового телевізійного мовлення ефірними, кабельними та супутниковими мережами;
DVB – (Digital Video Broadcasting) – сімейство стандартів цифрового телебачення;
VGA – (Video Graphics Array) – графічний інтерфейс та формат низької роздільної здатності;
VoD – (Video on Demand) - система індивідуальної доставки абоненту відео контенту телекомунікаційними мережами;
RAW – формат цифрового представлення мультимедійної інформації, що містить необроблені дані;
MPEG – сімейство стандартів відео кодування;
ДКП – дискретно-косинусне перетворення;
IF – (Image Fidelity) точність зображення;
SSIM – (structure similarity) – індекс структурної подібності;
PSNR – (Pulse Signal to Noise Ratio) пікове відношення сигнал/шум;
MSE – (mean square error) – середньоквадратичне відхилення (СКВ);
YCrCb – (y, chromatic red, chromatic blue) – яскравісно-хроматична колірна модель;
VBR – (variable bat-rate) – режим кодування з забезпеченням змінної інформаційної інтенсивності відео;
CBR – (constant bat-rate) – режим кодування з забезпеченням постійної інформаційної інтенсивності відео;
ABR – (constant bat-rate) – режим кодування з забезпеченням частково стабілізованої інформаційної інтенсивності відео;
GOP – (group of pictures) – група кадрів, структурна одиниця відео потоку;
M-JPEG – стандарт відео кодування;
M-JPEG2000 – стандарт відео кодування;
JPEG – технологія кодування зображень;
AVC – (advavced video coding) - технологія високоефективного кодування відео H.264;
HEVC – (high-efficiency video coding) — технологія високоефективного кодування відео H.265.

ВСТУП

Одночасно з розвитком інформаційних та комунікаційних технологій, мережевої інфраструктури та кінцевих пристроїв, здійснюється розвиток, а у ряді випадків – видозмінюється формат взаємодії «людина-інформаційне середовище».

Інакше кажучи, сьогодні зазнають суттєвих змін:

- способи та шляхи пошуку інформації особою у мережевому середовищі;
- шляхи та особливості взаємодії між особами на базі інформаціно-комунікаціних мереж.

При цьому, такі зміни характеризуються:

- зростанням середнього обсягу даних, який транслюється у ході одного сеансу взаємодії за напрямком «людина-інформаційний мережевий ресурс», або між двома чи більше учасниками взаємодії з використанням мережевого середовища;
- збільшенням середнього часу, який витрачається щоденно на перегляд інформаційних ресурсів або віддалені комунікації;
- збільшенням ролі мультимедійного, а першочергово - відеоконтенту на рівні майже усіх існуючих мережевих сервісів та мережевих ресурсів.

Наслідком зазначених змін є суттєва перевантаженість як мережевої інфраструктури, так і ресурсів пропускної спроможності мереж, що створює передумови до падіння якості послуг, що надаються користувачам на базі мережі.

Беручи до уваги те, що саме відеотрафіку відповідають найвищі рівні інформаційної інтенсивності, становиться зрозуміло, що саме ріст обсягу відео найбільш суттєвим чином зумовлює зростання навантаження на мережу.

У такій ситуації найбільш уразливими є:

- сервіси передавання даних критичного, інтерактивного та об’ємного типів;
- територіально розподілені технологічні системи, які передбачають або цілком, або на рівні окремих їхніх сегментів; функціонування у реальному часі;

- системи VoIP;
- відеосервіси VoD та інтерактивного типу та ін.

Означена проблематика є системною та на сьогоднішній час існує 2 ключових підходи до її вирішення, зокрема:

1. Менеджмент трафіку.

Даний підхід, реалізований у рамках системи QoS та з використанням її механізмів, дозволяє здійснювати управління інформаційними потоками для скорочення рівня навантаження на вузлах мережі шляхом їх перерозподілу.

Недоліком цього підходу є його ефективність в умовах, коли сумарне навантаження не перевищує деякого певного рівня, а також те, що менеджмент трафіку не здатен скоротити мережеве навантаження.

2. Обробка найбільш «важкого» трафіку – трафіку відео на рівні джерел.

У рамках цього підходу інформаційна інтенсивність відео багаторазово скорочується за рахунок технологій кодування. У підсумку цього отримується відеоінформаційний потік, параметри якого дозволяють здійснювати його трансляцію мережею зі зниженим ризиком викликати її перевантаження.

Водночас, для подальшого ефективного застосування даного підходу має бути вирішено ряд завдань, таких, як:

- забезпечення можливості кодування великих обсягів відео у реальному часі;
- створення умов, у яких можливий подальший ріст коефіцієнту стиснення без зменшення якості реконструйованого відео у точці прийому.

Існування означених завдань пояснюється:

- постійним стабільним ростом обсягу відеоінформації;
- посиленням вимог до якості відновленого відео як з боку користувачів, так і з позиції можливості ефективного використання засобів інтелектуального аналізу відеоінформації.

Зрозуміло, що рішення перелічених завдань необхідно будувати, приймаючи до уваги наступне:

- суттєве збільшення складності кодеку є неприпустимим, оскільки його застосовуваність при цьому обмежуватиметься лише високопотужними системами;

- найбільш прийнятними шляхами рішення означених завдань є удосконалення існуючих кодеків, що пов’язано зі складними економічними та технологічними аспектами;
- реалізація підходів до удосконалення існуючих кодеків має бути ефективною для переліку технологій, не обмежуючись однією.

1 ЗАСТОСУВАННЯ ВІДЕОІНФІОРМАЦІЇ У СУЧАСНОМУ СВІТІ ТА КЛЮЧОВІ ПІДХОДИ ДО ЇЇ ОБРОБКИ НА РІВНІ КІНЦЕВИХ ВУЗЛІВ

1.1 Сьогод’нішня роль відеоінформації у суспільстві

З точки зору користувача мережевих сервісів важко уявити галузь людської діяльності, де на сьогодні не отримали розвитку різноманітні відеоінформаційні послуги. Їхня широка застосовуваність сьогодні значною мірою впливає на формування самого суспільства, міжосібних зв’язків у ньому та зв’язків між тими чи іншими групами у соціумі.

З самого початку вектор розвитку цього класу мережевих сервісів зумовлювався двома головними чинниками – потребою у відповідних послугах та поточними можливостями існуючої технологічної бази для їх утілення.

У підсумку, зараз відеосервіси тією чи іншою мірою утілені у таких галузях, як (рис.1.1):

- зв'язок;
- освіта та навчання;
- галузь розваг;
- засоби масової інформації;
- системи безпеки;
- виробництво;
- smart-системи;
- дослідницька галузь та ін.

Без перебільшення можно зазначити, що на сьогодні відео є найбільш поширеним типом динах, що надсилається мережею.

Так як свідчать результати огляду [1, 2], доля відеоінформації у світі сьогодні у 4 рази перевищує аудіо, технологічний трафік, файлообмін, соціальний трафік, E-mail, та чати разом. При цьому, надалі така пропорція буде змінюватися у бік росту частки відео.

Окремо слід відмітити значну роль відеоінформації у найбільш прогресивних технологічних концепціях, що стосуються побудови smart-простору, робототехніки та нових засобів комунікацій.

Слід зазначити, що за останні роки суттєво збільшилася аудиторія осіб, що користуються відеоінформаційними засобами на постійній основі. Цьому сприяли як раціональні чинники, зокрема, згадані вище потреби та

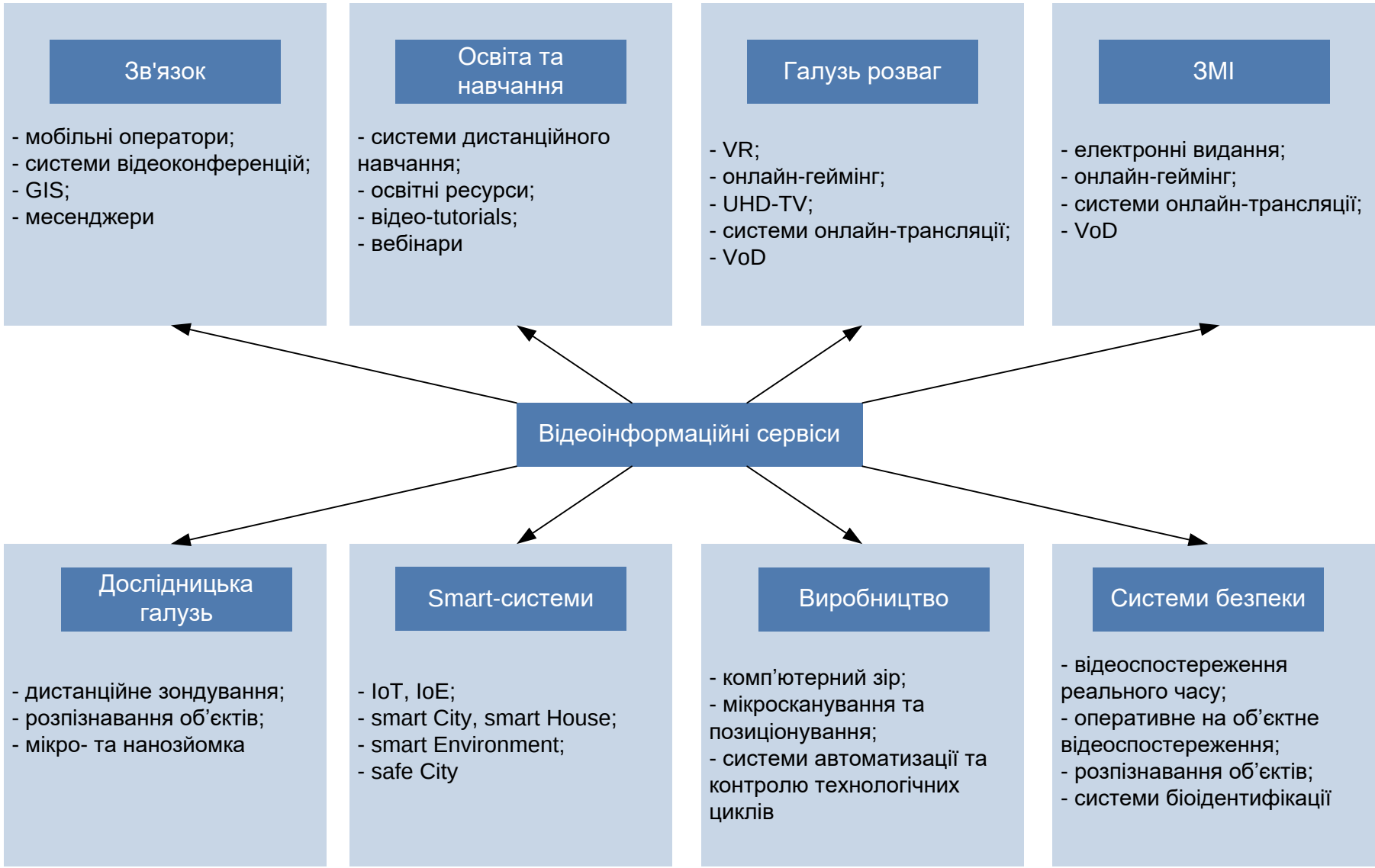


Рисунок 1.1 – Галузі застосування відеоінформації

технологічні можливості а також зовнішні фактори. Так, свого часу суттєвий поштовх у застосовуваності мережеві відеосервіси отримали у наслідок поширення пандемії Covid-19, що зумовило перехід багатьох сфер діяльності у віддалений режим (навчання, електронна торгівля, діяльність, що передбачає безпосереднє спілкування взагалі) з використанням засобів віддаленої комунікації. У таких умовах, з одного боку, суттєво зросла аудиторія мережевих платформ а вони самі зазнали розвитку.

З іншого боку, це сприяло виявленню багатьох проблемних аспектів як у функціонуванні самих таких платформ, так і концептуальних недоліків стосовно підходів до обробки відеоінформації, яка надсилається у мережу [3,4].

При цьому, якщо перша група проблемних аспектів, що стосуються безпосередньо мережевих платформ, може бути скорочена або взагалі усунена (застосуванням більш ефективного регламенту, нарощуванням потужності серверної частини тощо), то недоліки підходів до обробки відео потребують детального аналізу бачення процесу кодування, у сутності, покрокового опрацювання усього процесу для виявлення тих недоліків, існування яких сьогодні зменшує продуктивність функціонування відеоінформаційних сервісів.

Для того, щоб мати можливість локалізувати такі недоліки, розглянемо особливості існуючих кодеків та поширені сьогодні підходи до обробки відеоінформації.

1.2 Огляд ряду існуючих технологій кодування відео

Найбільш масовими та найчастіше використовуваними на сьогодні є технології, які відносяться до сімейства MPEG [5].

1.2.1 MPEG-1

Початок впровадження стандарту MPEG-1 у 1996 році створив передумови для широкого виходу відеоінформації у цифровий формат – спочатку для зберігання у файлосховищах та розповсюдження на носіях. Так, MPEG-1 початково розроблявся як кодек саме для VideoCD [6].

При цьому, передбачалася можливість побудови базового кодованого потоку відеокадрів роздільної здатності 352x240 з інформаційною інтенсивністю до 1,5 Мбіт/с.

Окрім базового показника роздільної здатності, існувала можливість кодувати відеоряд будь-якого формату до 4095x4095 включно. Даний факт пояснює причину орієнтованості MPEG-1 на оптичні носії – середня пропускна здатність мережевих каналів того часу не дозволяла здійснювати ефективні трансляцію. Разом з тим, згаданий стандарт не набув свого часу широкого розповсюдження. Це зумовлювалося наступним:

- існуючі обмеження на трансляцію відео мережею;
- суттєво обмежені можливості відеоадаптерів;
- поява пристроїв перегляду, роздільна здатність яких перевищувала VGA-формат (рис.1.2);
- переважна більшість відео записів здійснювалася у форматі 352x240.

У підсумку, частина користувачів не мала у своєму розпорядження достатніх обчислювальних спроможностей апаратної платформи.

При цьому, іншу частину користувачів не влаштовувала візуальна якість відеоряду низької роздільної здатності при відображенні його на дисплеї, роздільна здатність якого у кілька разів перевищувала формат відео.

Певний компроміс тут було досягнуто з впровадженням стандарту MPEG-2.

1.2.2 MPEG-2

Новий стандарт MPEG-2 надавав ряд переваг відносно MPEG-1, зокрема [7, 8]:

- вищий рівень K_{enc} коефіцієнту стиснення за тих же умов кодування;
- гнучка сітка стандартизованих роздільних здатностей відео;
- існування форматів 4:3 та 16:9;
- режими обробки progressive та interlaced (прогресивний та черезрядковий).

Так, на базі MPEG-2 забезпечувалась можливість кодування відео з роздільною здатністю 352x240 з інформаційною інтенсивністю до 0,384 Мбіт/с при 24 кадрах за секунду. Це створювало умови для можливості трансляції відео мережею.

Рисунок 1.2 - Приклад кадру відео роздільної здатності 352x240, кодованого на базі MPEG-1, на випадок його відображення на моніторі з роздільною здатністю 1024x768

Інакше кажучи, технологія MPEG-2 давала змогу досягти вищих рівнів стиснення відеоданих відносно MPEG-1, за умов збереження рівня якості реконструйованих даних незмінним.

Окрім описаних переваг MPEG-2 щодо передуючого стандарту стиснення, дана технологія містить у собі також протокол MPEG-TS, що виконує роль транспорту для відео та аудіоданих [9]. На базі MPEG-TS здійснюється мультиплексування пакетів аудіо та відеоданих, а також синхронізується їх вихід і надає можливості щодо корекції помилок таких транспортів, як DVB чи ATSC.

Дана закономірність пояснює те, що у таких умовах згодом виникли перші сервіси VoD, для яких спочатку MPEG-2 стала домінуючою технологією.

1.2.3 MPEG-4 та рекомендації H.26*

У свою чергу, широке впровадження технології MPEG-4 зумовило значне збільшення обсягу відеоінформації, яка надсилається мережею [1, 10].

Це пояснюється тим, що стандарт наслідував усі переваги, властиві MPEG-2, маючи при цьому ряд принципово нових функціональних можливостей. Так, наприклад, серед нових можливостей, які реалізовано у рамках MPEG-4 Part.10/H.264/AVC присутні такі [10, 11]:

- багатокадрове прогнозування;
- суттєво гнучкіше застосовувати попередньо кодовані кадри потоку як базові. Зокрема, передбачено можливість побудови до 32 посилань у потоці на інші кадры, тоді як попередні стандарти надавали змогу створювати щонайбільше 2 посилання (для кадрів B-типу). Такий підхід дозволяє кодеку для компенсації руху здійснювати вибір серед значної кількості зображень що, у свою чергу, збільшує продуктивність кодування;
- можливість виконувати процедуру компенсації руху на рівні блоків змінного розміру, а саме – 4x4, 8x8 та 16x16, за рахунок чого забезпечується можливість вибору ділянок руху значно точніше, ніж з використанням передуючих стандартів;
- використання алгоритмів екстраполяції за межі кадру для випадків існування векторів руху, які ведуть за межі видимої зони відеосцени;
- можливість виконувати компенсацію руху, забезпечуючи при цьому точність до чверті пікселя (Qpel) у випадку компонент яскравості і до $\frac{1}{8}$ пікселя для колірних компонент;
- використання гексагональної фильтрації компонент яскравості для забезпечення прогнозування з точністю до $\frac{1}{2}$ пікселя, що спрямовано на забезпечення чіткості кадру;
- застосування зваженого прогнозування, за рахунок чого досягається можливість виконувати зміщення та масштабування ділянки кадру після виконання процедури компенсації руху на величины, які встановлює кодек. Такий підхід сприяє збільшенню ефективності кодування просторових та часових ділянок відео потоку, де має місце зміна яскравості (наприклад, затемнення чи проявлення кадру або його частини);
- нові підходи щодо побудови процесу просторового прогнозування;
- можливість кодування відеокадрів на рівні блоків 4x4 та 8x8 пікселів;

- можливість кодування макроблоків без втрат;
- застосування деблокуючи фільтрів, що спрямовані на усунення артефактів блочності, властивих ДКП-базованим методам кодування;
- використання контекстно-залежного адаптивного двійкового арифметичного кодування на етапі кодування без втрат, як альтернатива методу Хафмана.

У підсумку, відносно передуючих технологій кодування, за одних і тих же умов кодек MPEG-4 дозволяє досягти досить високих показників K_{enc}, забезпечуючи при цьому якість відновлених даних.

Сьогодні, беручи до уваги усі перелічені переваги, стандарти MPEG-4 Part 10 та рекомендація H.264/AVC є одними з найчастіше застосовуваних у світі. Так, зараз кодек H.264 за замовчуванням використовується для:

- камер більшості мобільних пристроїв;
- відеокамер, у т.ч. action-камер;
- більшості пристроїв фото і відеофіксації;
- зберігання даних у файлосховищах.

Водночас, дещо менший відсоток пристроїв з перелічених категорій сьогодні базується на використанні кодеку H.265/HEVC, що, у сутності, може розглядатися як розвиток рекомендації H.264. Застосування кодеку H.265 забезпечує вищий рівень якості реконструйованого відео порівняно з H.264 при одних і тих же K_{enc}. На сьогодні рекомендація H.265/HEVC дещо менш поширена, ніж H.264, оскільки кодек має складнішу реалізацію, що накладає додаткові вимоги щодо обчислювальної потужності для апаратних платформ, на базі яких планується його використання.

1.2.4 Сімейство M-JPEG

До даної групи входить ряд технологій, що розглядають відеопотік, як сукупність незалежних між собою кадрів. При цьому, кодування кожного з них здійснюється за ідеологією JPEG з рядом модифікацій. Ключовими відмінностями даної технології від MPEG є [12]:

- відсутність класифікації кадрів на I, P та в типи та інших механізмів усунення часової надмірності;
- у середньому у 3-7 разів вища інформаційна інтенсивність кодованого відео потоку ніж для кодеків групи MPEG;
- вища середня якість довільних стоп-кадрів з потоку.

Остання з перелічених особливостей є причиною, за якою пристрої фото та відеофіксації на базі M-JPEG зараз використовуються профільними відомостями у місцях з великим скупченням людей (місця, популярні для відвідування, транспортні вузли тощо).

На засадах, аналогічних M-JPEG, функціонує кодек M-JPEG2000, де у якості ключового алгоритму кодування використовується JPEG2000. Відмінністю M-JPEG2000 є вищі рівні K_{enc} за умов сталої якості відео, та вища складність реалізації.

1.2.5 Фрактальні методи кодування

Кодеки, що базуються на принципах фрактального стиснення, принципово відрізняються від усіх, зазначених вище. В їх основі – процеси пошуку самоподібних структур зображення та відповідних афінних перетворень [8]. При цьому самоподібними структурами зображення є такі його елементарні ділянки, на базі яких може бути побудовано інші його частини шляхом:

- зміни масштабу самоподібних структур;
- зміни координат їх розміщення у зображенні;
- зміни куту нахилу самоподібної структури.

У свою чергу, афінне перетворення відповідає за виконання перелічених процесів.

Теоретично на базі фрактальних кодеків може бути досягнуто коефіцієнтів стиснення $K_{enc} \approx 2000$.

У той же час, фактичний його рівень суттєво залежний від особливостей змісту конкретного зображення та не забезпечується сталим, знаходячись у діапазоні $K_{enc} = (0{,}5; 2000)$.

Тобто, у результаті кодування розмір зображення може навіть збільшитися. Таке можливе тоді, коли для середнього розміру $\overline{Z}$ самоподібної структури справедливо:

$$\overline{Z} \to 1 \qquad (1.1)$$

тобто, тоді, коли розмір наближається до пікселя.

Таку залежність, як показано виразом (1.1), частково може бути усунено за рахунок корегування порогу допущень, що, у свою чергу, умовно збільшує величину $\overline{Z}$, погіршуючи при цьому якість відновленого зображення.

Ще однією ключовою відмінністю між фрактальними методами кодування та MPEG-орієнтованими методами є чітко виражена їхня асиметричність, що може досягати порядку 1:10000.

Так, час кодування одного зображення з використанням фрактального стиснення може сягати 1-2 секунд, у той час як декодування – порядку 0,001-0,0001 секунди.

У підсумку, існуюча специфіка кодеків даного типу зумовила галузі їх застосування – фрактальні кодеки неможливо застосовувати для обробки відео у реальному часі, натомість на їх базі може бути створено бібліотеки відеофайлів, побудовано сервіси VoD тощо.

1.3 Ключове завдання технологій кодування відеоінформації

Будь-які технології кодування відеоінформації слугують для скорочення інформаційної інтенсивності R відеопотоку.

Це зумовлено тим, що величина R відео потоку, який попередньо не було кодовано (т.з. RAW-потік), не дозволяє забезпечити його трансляцію існуючими каналами [11].

Наприклад, якщо мова йде про відео роздільної здатності 1920x1080, для його трансляції у реальному часі необхідно забезпечити пропускну здатність на рівні не нижче, ніж 600 Мбіт/с.

Зрозуміло, що трансляція такого RAW-потоку не можлива ні зараз, ні у подальшому з використанням, наприклад, технологій 5 покоління, оскільки тут мова йде про єдиний потік [3].

У той же час, відеоінформаційний потік, кодований на базі H.264, потребує для трансляції пропускну здатність не більш, ніж 3-4 Мбіт/с за найгірших умов.

Така розбіжність між інформаційною інтенсивністю кодованого та RAW-потоків викликана існуванням високого рівня надмірності відеоданих, найвищого порівняно з тими, що відповідають іншим типам даних.

Зокрема, зменшується вплив надмірностей двох головних класів – просторової та часової [11, 13].

Тут під ***часовою надмірністю*** мається на увазі властивість відео потоку містити у собі статичні та динамічні ділянки відеосцени.

Відтак, статичні ділянки повністю чи частково можуть дублюватися у певній сукупності послідовних кадрів потоку, що зумовлює невиправданий ріст R інформаційної інтенсивності (рис.1.3).

а)

б)

Рисунок 1.3 – Приклади фрагменту відеоряду, відносно якого було виконано процедури усунення часової надмірності а) та відеоряду, де часову надмірність скорочено не було

У свою чергу, алгоритми скорочення часової надмірності є у складі технологій сімейства MPEG та ряду технологій, що побудовані за фрактальними принципами. На їх базі сьогодні досягається скорочення інформаційної інтенсивності у середньому у 2,5-5 разів (для рекомендацій H.26*).

При цьому, найбільш суттєвий вплив на підсумковий рівень інформаційної інтенсивності відео здійснює ***просторова надмірність***, що, у свою чергу, розподіляється на такі базові типи надмірностей, як:

- психовізуальна;
- семантична;
- ймовірнісно-статистична;
- комбінаторна;
- структурна.

Поширені зараз технології кодування орієнтуються, головним чином, на скорочення ймовірнісно-статистичної, частково-структурної та психовізуальної надмірностей.

Так, психовізуальна зумовлена специфікою людського зору, та його

несприйняттям ряду класів викривлень зображень.

При цьому, у ході скорочення даного типу надмірності зазвичай втрачається деяка частина інформації про яскравість та колірність відео зображення. Разом з тим, візуальна якість відеокадрів при цьому не знижується.

У сьогоднішньому технологічному базисі скорочення надмірності психовізуального типу забезпечує найбільшу ступі зменшення інформаційної інтенсивності відео, ніж для випадку надмірностей будь-якого іншого типу.

Власне, самі алгоритми усунення психовізуальної надмірності, як вже було зазначено, базуються на особливостях людського зору, серед яких головними є:

1) особливості сприйняття рівнів яскравості. Так, зір людини значно чутливіший до яскравісних градацій зображень, ніж градацій колірних відтінків;

2) особливості сприйняття частотних компонент. Переважна більшість семантично значимої інформації містять у собі низькочастотні компоненти зображення. При цьому, зір людини добре розрізняє якраз низькочастотні складові і значно менше – високочастотні, які не несуть у собі суттєвої інформації.

Головними інструментами скорочення психовізуальної надмірності для MPEG, M-JPEG та M-JPEG2000 при цьому є [8, 11, 13, 14]:

- інструментарій вибору формату колірного опису, що у межах кожного окремого відеокадру здатен на 3/4 скоротити кількість інформації про колірність;
- квантування, у процесі якого низькочастотні складові зазнають незначних змін, тоді як високочастотні суттєвим чином скорочуються, або взагалі обнуляються.

У свою чергу, існування надмірності ймовірнісно-статистичного типу викликана присутністю зв'язків ймовірнісно-статистичного характеру, які пов’язують елементи просторового опису кадру та потоку на різних рівнях представлення (внутрішньо кадрові, такі, що пов’язують колірні компоненти, між кадрами у групі тощо).

Скорочення внутрішньо кадрових надмірності ймовірнісно-статистичного типу здійснюється застосуванням кодування без втрат на останніх етапах перетворення інформації/

При цьому, фрактальні методи орієнтуються на усунення структурної надмірності у просторовій площині.

Таким чином, головне завдання процесу кодування полягає у стисненні відео потоку шляхом скорочення різних типів надмірності з забезпеченням максимально можливого коефіцієнту K_{enc} стиснення. Водночас, у наслідок скорочення надмірності має забезпечуватися якість реконструйованих даних не нижче потрібного рівня.

При цьому, для оцінки якості відео може бути застосовано показник D рівня внесеної помилки який, у свою чергу, може бути виміряно [11, 15]:

- середньоквадратичним відхиленням (MSE);
- піковим видношенням сигнал-шум;
- рівнем точності зображення (IF);
- абсолютним відхиленням (AE);
- рівнем структурної подібності (SSIM).

Тоді ключова вимога то технології кодування може бути подана наступним виразом:

$$K_{enc} \uparrow | D \le D', \tag{1.2}$$

де D' - рівень внесеної помилки, допустимий за даних умов.

1.4 Недоліки існуючих відеокодеків

Процес стиснення відеопотоку у рамках широко використовуваних зараз підходів являє собою застосування тих чи інших параметрів кодування для забезпечення виконання умов (1.2).

На сьогодні у межах схеми MPEG, а саме – рекомендацій H.26* параметрами, які можуть налаштовуватися, є крок ν_q квантування та формат $Y:Cr:Cb$ (модель) колірної субдискретизації. Ще один з етапів обробки відео – статистичне кодування – не має доступних для налаштування опцій. Відповідно, за замовчуванням справедливою є наступна функціональна залежність:

$$R, D = \psi(K_{enc}) = \varphi(\nu_q; Y:Cr:Cb). \tag{1.3}$$

Разом з тим, величина R відео потоку, який надсилається мережею, має встановлюватися виходячи з:

- фактичної величини W пропускної знадності мережі;
- рівня D допустимої помилки.

Отже, відповідним чином має змінюватися коефіцієнт K_{enc} стиснення. Це еквівалентно виразу:

$$\begin{cases} R = \varphi_1(K_{enc}) = \psi_1(D) \,|\, R < W; \\ R = \varphi_2(K_{enc}) = \psi_2(W) \,|\, R \geq W, \end{cases} \tag{1.4}$$

де φ_1, ψ_1, φ_2 та ψ_2 - функціонали, що пов'язують величини R та K_{enc} з W та D.

Водночас, виконання нерівності (1.4) на сьогодні не гарантується, оскільки існуючі технології кодування сьогодні не мають стандартизованих механізмів урахування величин W та D при визначенні необхідного рівня K_{enc} коефіцієнта стиснення.

Наслідком цього є втрати пакетів даних у процесі транслювання та ріст рівня затримки їх доставки, що у підсумку веде до порушення якості відеоряду, або взагалі унеможливлює його надсилання.

Окрім зазначеного, аналіз виразу (1.3) показує, що існуюча залежність потенційно створює умови для росту величини D, а також можливого безпідставного збільшення величини R.

Це пояснюється тим, що поширені технології кодування не мають чіткого інструментарію на рівні стандарту, який встановлював би величину K_{enc} коефіцієнту стиснення відповідно не лише до необхідного рівня D, але також брав до уваги множину особливостей змісту оброблюваних фрагментів відеопотоку, як показує вираз:

$$K_{enc}, D = \varphi(\nu_q; Y : Cr : Cb; \{G\}; \varepsilon) \tag{1.5}$$

де {G} - множина статистичних, семантичних та психовізуальних особливостей оброблюваного фрагменту;

ε - рівень поточної декомпозиції відеоінформаційного потоку, на якому виконується кодування.

Отже, необхідно дослідити можливість виконання умов (1.5), що потенційно сприяють забезпеченню встановлення раціональних величин R та D у ході кодування.

2 ОСОБЛИВОСТІ ОБРОБКИ ВІДЕОІНФОРМАЦІЙНИХ ПОТОКІВ У РАМКАХ ІСНУЮЧОГО ТЕХНОЛОГІЧНОГО БАЗИСУ

2.1 Ключові вимоги до процесу кодування динамічних відеоінформаціних ресурсів

Як зазначалося вище, кодований відеопотік має відповідати вимогам, наведеним виразом (1.2). Разом з тим, існує ряд ключових вимог до ефективного процесу кодування динамічних відеоінформаціних ресурсів, серед яких першочерговими є [15]:

- вимоги до оперативності кодування;
- вимоги до точності обробки;
- вимоги до ресурсоємності кодеку.

Тут під ***оперативністю кодування*** розуміється властивість кодеку здійснювати обробку надходячого відеоряду за час Δt, величина якого наближається до мінімально можливого t_{min} у поточних умовах, які ураховують доступний ліміт P обчислювальних ресурсів системи, множину {A} особливостей змісту відеоряду та продуктивність W самого кодеку. Дана залежність аналітично може бути описаною наступним виразом:

$$\begin{cases} \Delta t \to t_{min}; \\ \Delta t = \varphi(\{A\}; P; W). \end{cases} \quad (2.1)$$

Існування даної вимоги зумовлене, з одного боку, тим, що в умовах, коли відеоконтент становить понад 80% загального обсягу мережевого трафіку, несвоєчасна обробка відео веде до порушення стандартизованих умов трансляції трафіку інших типів, що, у свою чергу, спричинює функціонування багатьох сервісів у аварійному режимі. За великим рахунком, некритичними до часу кодування є лише процеси підготовки відеоінформації до тривалого зберігання. Для цього випадку значно важливішим є час декодування.

З іншого боку, на сьогодні значний відсоток відеотрафіку належить сервісам, що функціонують у режимі реального часу:

- системи відео нагляду реального часу;

- платформи відео конференцій;
- інтерактивні відеосервіси;
- VR-системи;
- системи відеозв’язку.

Зрозуміло, що для перелічених систем та сервісів існування суттєвої затримки Δt є неприпустимим.

У свою чергу, ***точність обробки*** характеризує властивість кодеку забезпечувати раціональне розподілення інформаційної інтенсивності R між структурними одиницями динамічного відеопотоку.

При цьому, передбачається, що кожна структурна одиниця потоку у наслідок кодування повинна описуватися такою кількістю біт, яка, з одного боку, жодним чином не веде до скорочення її семантичного змісту S нижче критичного рівня S_{cr} за умови, що одночасно з цим забезпечується максимальна величина коефіцієнта $К_{enc}$ скорочення інформаційної інтенсивності.

Наближено коефіцієнтом зміни семантичного змісту S динамічного відеопотоку може вважатися показник помилки D реконструйованої інформації відносно вихідної.

Залежно від того, яка саме оцінка рівня D помилки у тому чи іншому випадку є першочерговою (візуальна чи об’єктивна), серед зазначених вище метрик оцінки якості (або помилки) може бути використано як ті, що характеризуються суттєвою розбіжність між математичною оцінкою та оцінкою, виконаною експертом, так і ті, що максимально збігаються за даним показником.

У першому випадку це стосується найбільш поширених зараз метрик – MSE, PSNR, IF та ін. Для даних метрик в окремих випадках об’єктивна оцінка може свідчити про достатній рівень якості відновленої відеоінформації, тоді як візуально відеоряд може бути зруйнованим, або навпаки.

Така залежність зумовлена тим, що традиційні метрики оцінювання якості відео та зображень є значно залежними від особливостей змісту.

У той же час, максимально наближеною за об’єктивною оцінкою до результатів оцінок, отриманих за участю експертів, є метрика SSIM. При цьому, за результатами неодноразових досліджень [16] встановлено відповідність таких оцінок для контенту довільного змісту.

Таким чином, ***точність обробки*** можемо вважати поняттям, що є тотожним поняттю раціонального розподілу інформаційної інтенсивності у межах відеопотоку.

У випадку, коли розглядаються окремі кадри, точність Ф обробки може вказувати на:

- ступінь відповідності Λ семантичного змісту $S(F)_{x,y}$ кодованого фрагменту відеопотоку кількості біт $R(F)_{x,y}$, що застосовано для його представлення;
- рівень D внесеної похибки у результаті кодування Така залежність між рівнем внесеної помилки D, величиною коефіцієнта стиснення та точністю;
- рівень U остаточної надмірності даних, які надсилаються у мережу;

Тобто, точність обробки можемо інтерпретувати як функціонал від величин Λ та D, що еквівалентно виразу:

$$\Phi = \varphi(\Lambda; D; U) \tag{2.2}$$

При цьому, підвищення точності обробки відеоряду потенційно здатне вирішити питання скорочення інформаційної інтенсивності з одночасним забезпеченням візуальної та об'єктивної якості реконструйованих даних, як показує вираз:

$$\Phi\uparrow \ \rightarrow \ R\downarrow | D \approx \text{const}. \tag{2.3}$$

При цьому, поняття ***ресурсоємності кодеку*** може інтерпретуватися як об'єм Q операцій обчислень, необхідних для формування коду найменшої структурної одиниці динамічного відеопотоку.

2.2 Специфічні особливості динамічних відеоінформаційних потоків

Трансляція відеоінформаційних потоків у мережі може здійснюватися у таких режимах, як [13, 15]:

- потоковый режим;
- пульсуючий режим.

При цьому, у потоковому режимі відеотрафік являє собою рівномірний потік даних, що характеризується інформаційною інтенсивністю, яка є незмінною у часі.

Ключова вимога до рівня інформаційної інтенсивності відео для цього режиму формування потоку є наступною:

$$R \to \text{const}. \tag{2.4}$$

Для формування такого трафіку відео використовується CBR (Constant Bit Rate)-режим кодування.

Трафік такого типу утворюється послідовністю пакетів, кожен з яких займає однакову кількість $R_{\text{раск}}$ біт. Такі пакети надсилаються один за одним через один и той же часовий інтервал ΔT (рис. 2.1).

Для цього випадку бітова інтенсивність R_{CBR} такого потоку CBR може бути визначена на базі наступного виразу:

$$R_{CBR} = \frac{R_{\text{раск}}}{\Delta T}. \tag{2.5}$$

Також у цьому випадку значення величини R_{CBR} потокового трафіка буде нижчим, ніж максимально допустиме значення, що визначається протоколом, на базі якого здійснюється передавання даних, що зумовлено наявністю интервалів між пакетами.

Безумовною перевагою відеотрафіку потокового типу є рівномірне та прогнозоване навантаження на мережу а також на мережеві вузли.

Це, у свою чергу, сприяє суттєвому скороченню ймовірності перевантаження мережі, оскільки, з одного боку, майже повністю виключає випадки недетермінованого лавиноподібного підвищення діючої інформаційної інтенсивності, а з іншого боку – дозволяє керувати наявними мережевими ресурсами, зокрема:

- резервувати канали;
- будувати альтернативні обхідні напрямки при формуванні таблиць маршрутизації;

- виділяти фіксований обсяг обчислювальних ресурсів для обробки відеоінформації на кінцевих пристроях, тощо.

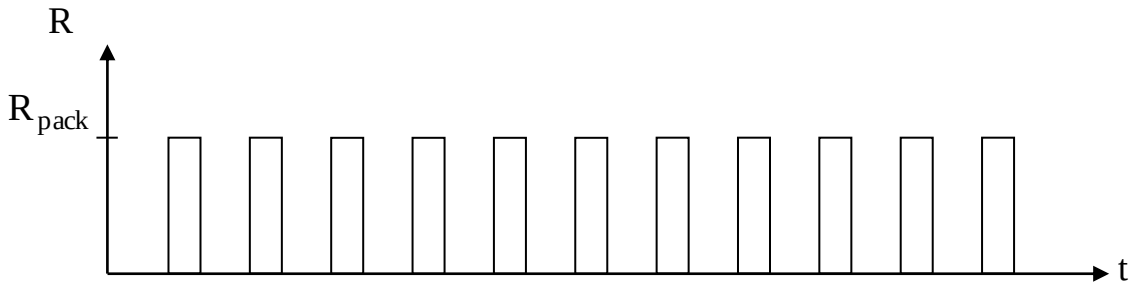


Рисунок 2.1 – Приклад потокового трафіку

Разом з тим, підходу до формування відеоінформації на засадах CBR властивий також ряд недоліків, серед яких головним є нераціональність використання пропускної здатності К мережі. Даний недолік зумовлений тим, що:

1. З одного боку, рівень інтенсивності R відеоінформаціних потоків початково є нерівномірним у часі, тобто:

$$R(t_1) \neq R(t_2) \ldots \neq R(t_n) \qquad (2.6)$$

У загальному випадку відеоінформаційний потік спрощено може розглядатися як певним чином організована сукупність окремих кодованих кадрів, що змінюють один одного у часі, як показує рисунок.2.2, а).

Для такого типу трафику існує, щонайменше, два чинника, які зумовлюють виникнення пульсацій рівня інформаційної інтенсивності, зокрема:

- суттєві відмінності особливостей внутрішнього змісту кадрів потоку, що надходять на обробку. У свою чергу, спричинювати наявність таких відмінностей можуть, зокрема, відео сцени, для яких характерною є висока динаміка (так звана часова неоднорідність), а також – просторова неоднорідність у межах кожного окремого кадру;

- у рамках існуючої сьогодні моделі кодування на засадах MPEG, реалізованої сьогодні для переважної більшості відеокодеків, передбачено

принцип формування відео потоку на основі кадрів трьох типів – опорних, або ключових I-кадрів, передбачених P-кадрів та двоспрямовано-передбачених B-кадрів. Кадри усіх даних типів надсилаються у деякій послідовності, що визначається стандартом кодування. При цьому, для представлення кожного з кадрів необхідна різна кількість біт R_F. Зокрема, найбільша кількість біт, та, відповідно, інформаційна інтенсивність відповідає I-кадрам. Загальний характер співвідношення інформаційних інтенсивностей кадрів різних типів показує наступне співвідношення [8, 11]:

$$R_F^{(I)} > R_F^{(P)} > R_F^{(B)} \qquad (2.7)$$

Зазначену виразом (2.7) залежність ілюструє також рис.2.2.б).

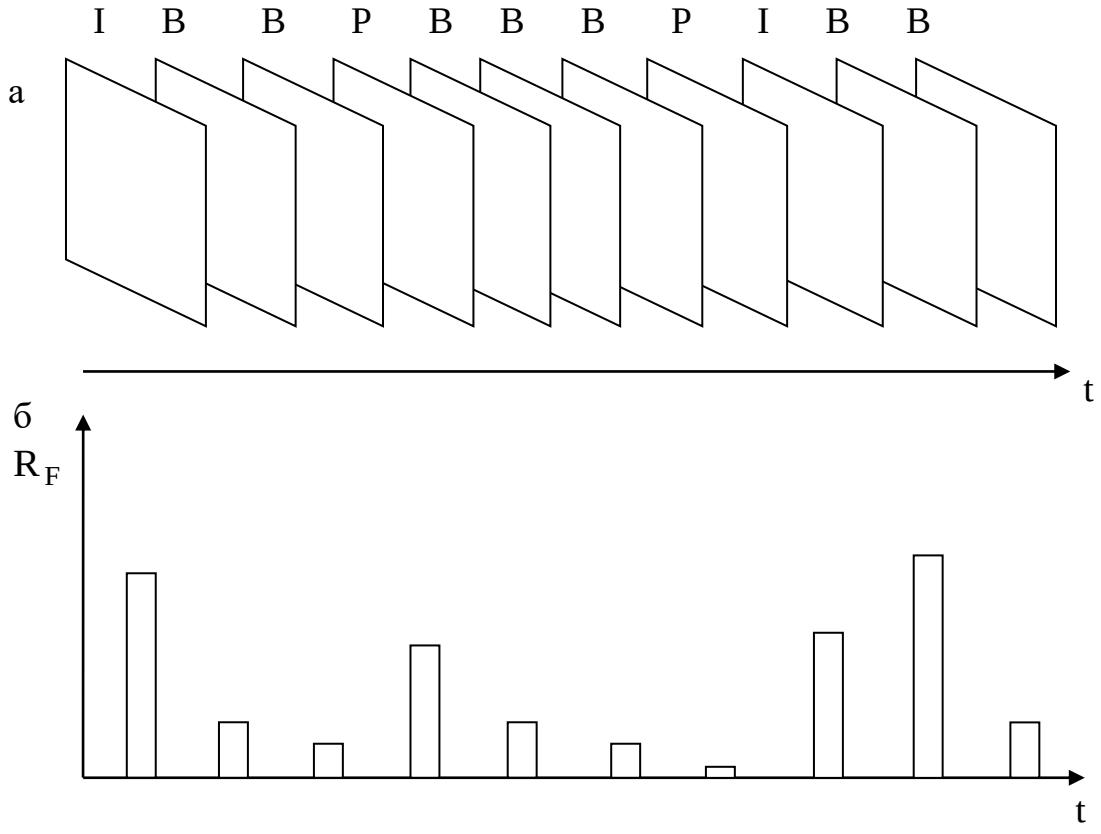


Рисунок 2.2 – а) Загальна структура відеоінформаційного потоку, утвореного у на базі MPEG, б) типовий приклад ррозподілу інтенсивності між окремими кадрами у відеопотоці

Також, окрім особливостей внутрішнього змісту окремих кадрів, та застосовуваного методу обробки відеоданих, суттєвим чином величину пульсацій інформаційної інтенсивності R у ході передавання відеопотоку здійснюють вплив специфіка технологічних рішень та задіяних параметрів режиму кодоутворення у рамках застосовуваної базової технології.

2. Величина інформаційної інтенсивності R_{CBR} потоку має підтримуватися сталою. Таким чином, якщо у деякий довільний момент часу t значення інформаційної інтенсивності зменшується, тобто $R(t) < R_{CBR}$, у цьому випадку необхідно збільшити її величину до $R(t) \approx R_{CBR}$. Для цього у загальному випадку застосовуються такі методи, як:

- зміна коефіцієнту $К_{enc}$ стиснення (зменшення його величини для наступного часового інтервалу $(t+1)$);

- додавання до потоку даних, що надходить від джерела, так званих пакетів-заповнювачів. На сьогодні даний підхід отримав реалізацію, наприклад, для DVB-TV.

Розглянемо детальніше кожен з підходів до утворення CBR-потоку відео.

2.3 Проблематика формування CBR-потоків відео

Так, у підсумку, беручи до уваги неоднорідний характер інформаційної інтенсивності відео вже на рівні джерела, стає зрозумілим, що залежно від особливостей змісту відеоряду потік буде характеризуватися тим чи іншим коефіцієнтом $К_{pulse}$ пульсації, при цьому $К_{pulse} \neq 0$. У свою чергу, даний показник розраховується для деякого часового інтервалу $\Delta T = [t; t + \vartheta]$, як:

$$К_{pulse} = \frac{R(\Delta T)_{max} - R(\Delta T)_{min}}{2R(\Delta T)_{med}}, \qquad (2.8)$$

де $R(\Delta T)_{max}$ та $R(\Delta T)_{min}$ - величини максимальної та мінімальної інформаційної інтенсивності потоку відео за інтервал, що підлягає спостереженню;

$R(\Delta T)_{med}$ - середнє значення інформаційної інтенсивності потоку відео за інтервал.

Отже, для того, щоб утворився CBR-потік, у виразі (*.7) необхідно або збільшити величину $R(\Delta T)_{min}$ до значення $R(\Delta T)_{max}$, або навпаки, забезпечити умови скорочення $R(\Delta T)_{max}$ до $R(\Delta T)_{min}$.

У першому випадку для того, щоб зрівняти R_{max} та R_{min}, може здійснюватися контроль коефіцієтна K_{enc} стиснення відповідно до наступного принципу:

$$K(t+1)_{enc} := K(t)_{enc} - \eta \mid R(t) < R(\Delta T)_{max}, \qquad (2.9)$$

де η - величина зміни коефіцієнта стиснення.

Тобто, якщо у поточний момент часу величина $R(t)$ інформаційної інтенсивності є меншою, ніж максимальна $R(\Delta T)_{max}$ за інтервал ΔT, для кодування деякого обсягу відеоданих у наступний момент часу $(t+1)$ буде задіяно коефіцієнт стиснення, K_{enc} зменшений відносно поточного на деяке значення η.

Такий підхід може бути реалізовано лише за умови, що потік відеокадрів розглядається як сукупність локальних ділянок $\Delta T = [t; t+\vartheta]$ у часовому просторі (рис.2.3). У загальному випадку розмір кожної такої ділянки можу бути відмінним від інших, тобто, $\Delta T \neq (\Delta T - \mu) \neq (\Delta T + 1) \neq (\Delta T + \psi)$. При цьому, множина біт, що утворюється у наслідок кодування такого часового інтервалу, формує пакет даних, що далі надсилається мережею.

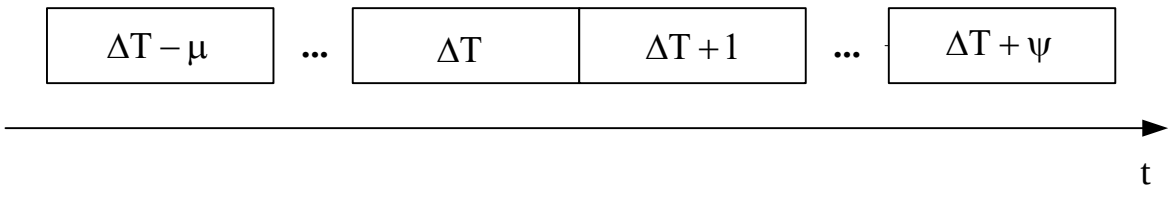


Рисунок 2.3 – Побудова відео потоку на засадах CBR на рівні сукупності локальних часових ділянок

Разом з тим, підхід на базі управління коефіцієнтом K_{enc} стиснення є суттєво обмеженим з тієї причини, що для ряду локальних часових ділянок ΔT забезпечити умови, у яких $R(\Delta T)_{min} \rightarrow R(\Delta T)_{max}$, неможливо, коли на обробку за інтервал ΔT надходить деяка частина фрагментів відеопотоку,

яким відповідають незначні рівні інформаційної інтенсивності. Це характерно для випадків, коли:

- мають оброблятися кадри P та B-типів, що апріорі можуть мати низькі рівні величин $R_F^{(P)}$ та $R_F^{(B)}$ (рис.2.2);

- внутрішній зміст кадрів будь-якого типу дозволяє здійснювати їх опис невеликим обсягом біт.

Зрозуміло, що за таких умов справедливим є співвідношення:

$$K(t+1)_{enc} := K(t)_{enc} - \eta_{max} = K_{enc}^{(min)} \quad \nrightarrow R(\Delta T)_{min} \rightarrow R(\Delta T)_{max} \qquad (2.10)$$

Отже, у зазначених випадках, динамічна зміна коефіцієнта K_{enc} стиснення не гарантує забезпечення виконання умов (2.3). У зв'язку з цим для корегування рівня $R(t)$ інформаційної інтенсивності до значення $R(\Delta T)_{max}$ до послідовності кодованих даних додаються пакети-заповнювачі, як згадувалося раніше. Сутність збільшення рівня $R(t)$ у такий спосіб ілюструє приклад нормованої діаграми, зображеної на рисунку 2.4.

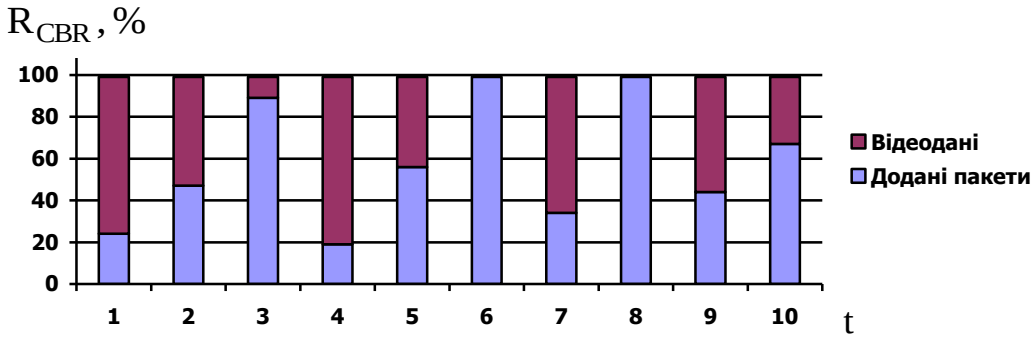


Рисунок 2.4 – Нормована діаграма розподілу інформаційної інтенсивності між пакетами відеоданих та заповнюючими пакетами у ході формування CBR-потоку

Тобто, як свідчить аналіз рисунку 2.4, у процесі застосування пакетів-заповнювачів для формування CBR-трафіку пропускна здатність мережі використовується нераціонально, оскільки при цьому суттєвий відсоток пакетів при цьому не несуть у собі інформаційного наповнення.

Отже, у підсумку можна зробити такі висновки:

- формування відеоінформаціних потоків може виконуватися на базі CBR та VBR підходів;
- відеоінформаційний потік з огляду на особливості обробки у схемі MPEG, а також існуючу динаміку змін відеосцен у часі, має пульсуючий характер;
- використання CBR-відеотрафіку створює умови, за яких ймовірність перевантаження мережі зменшується, оскільки при цьому мережеве навантаження може бути певною мірою прогнозованим а відтак – може бути виконано заходи (резервування ресурсів, розробка відповідних сценаріїв обробки трафіку тощо), спрямовані на її підтримку у керованих межах;
- реалізація CBR-підходу під час кодування відео веде до нераціонального використання мережевих ресурсів (оскільки частина навантаження при цьому створюється додатковими пакетами нульової інформативності) та нераціонального розподілу інтенсивності усередині відеопотоку.

З огляду на зазначене, на сьогодні переважна більшість відеоінформаціних потоків, що надсилаються мережею, кодуються відповідно до VBR-засад.

2.4 Умови раціонального розподілу інформаційної інтенсивності усередині кодованих відеокадрів

Розглянемо умови, за яких можливо збільшення точності обробки відеокадрів. Для цього спершу виконаємо аналіз двох відеокадрів, отриманих у межах одного потоку RAW (рис.2.5), тобто, без застосування стиснення.

Кожен з зазначених кадрів має роздільну здатність 1366x768 пікселів. При цьому у випадку, що розглядаються RAW-відеокадри, тобто, такі, що не зазнали процедури кодування, кількість біт $R(F)$ для опису кожного з них може бути визначено за виразом [11]:

$$R(F) = \frac{H \times W \times b}{1024}, \qquad (2.11)$$

де H та W - висота та ширина відеокадру у пікселях відповідно;

b - розрядність, або кількість байт для представлення одного пікселя.

У нашому випадку RAW-формат представлено 24-розрядною колірною палітрою. Тобто, b=3. Тоді, відповідно, $R(F) = \frac{1366 \times 768 \times 3}{1024} = 3{,}07$ МБ.

а)

б)

Рисунок 2.5 – Кадр1 а) та кадр 2 б), що належать одному RAW-потокові, MSE=0

Далі відеопотік підлягає кодуванню на базі H.264, у ході здійснюється стиснення кадрів з коефіцієнтом $K_{enc} = 60$. У результаті цього утворився масив I, B та P кадрів.

При цьому, кількість біт, що необхідні для представлення кадрів I-типу, що у кодованому потоці відповідають RAW-кадрам вихідних відеоданих, стає рівною близько 50 КБ.

Одержані таким чином I-кадри зображено рис.2.6.

а)

б)

Рисунок 2.6 – I кадр1, MSE=121207 а) та I-кадр 2, MSE=385728 б), що належать одному кодованому потокові

Як бачимо з рис. 2.5 та рис.2.6, у наслідок кодування з однаковими параметрами, спостерігається суттєве падіння візуальної якості кадру 2, тоді як візуально помітних змін у якості кадру 1 не виявлено. У свою чергу, для кадру 2 величина MSE складає 586728, а для кадру 1 – 121207, тобто, різниця є майже п’ятикратною.

Така розбіжність у значенням MSE для кадрів однакової роздільної здатності за умов однакової величини $К_{enc}$ пояснюється тим, що:

- MPEG-базовані технології кодування, у т.ч. і H.264, для обробки відео потоку на рівні окремих кадрів застосовують алгоритм JPEG разом з додатковими механізмами (наприклад, ДІКМ);

- алгоритм JPEG є контентно-залежним, тобто, стиснення різних ділянок кадрів за однакових умов веде до утворення кодованих ділянок, що відрізняються показниками якості.

Для дослідження даної закономірності далі виконаємо дослідження локалізації помилок, внесених за результатами кодування, у кожен з досліджуваних кадрів. Для цього може бути використано програмний засіб DiffImg [17]. Результати порівняння вихідних кадрів, та кадрів, одержаних у результаті кодування, а також відповідні карти розподілу помилок для кожного з досліджуваних кадрів наведено рис.2.7.

а)

б)

Рисунок 2.7 – Карти помилок, побудовані за результатами порівнянь вихідних та кодованих файлів - кадру 1 а), кадру 2 б)

Рисунок 2.7 показує, що у випадку кадру 1 максимум помилок сконцентровано у контурних зонах. У свою чергу, для кадру 2 спостерігається велика кількість помилково відновлених пікселів, що виражено як у зонах контурів, так і у зонах, що мають текстурне заповнення.

Саме ці відмінності – різний обсяг контурної інформації, а також присутність текстурних ділянок у кадрі 2 зумовили різний рівень візуальних спотворень та величин MSE.

Разом з тим, основна частина семантичного змісту у межах будь-якого зображення міститься саме у контурних ділянках.

Таким чином, з точки зору підвищення точності обробки відеопотоку, доцільним є підхід, у рамках якого обробка окремих фрагментів виконується з огляду на факт належності контурів у її складі, що потенційно сприяє виконанню умов (1.5).

Отже, надалі для реалізації даного підходу необхідно вирішити такі завдання, як:

- визначення механізму виокремлення контурних ділянок з кадру;
- вироблення рішення щодо рівня декомпозиції відео інформаційного потоку, на якому виконується обробка.

2.5 Висновки за розділом

За результатами викладеного матеріалу може бути зроблено висновки про наступне:

- одними з ключових показників, що характеризують ефективність процесу кодування динамічних відеоінформаціних ресурсів а відтак – застосованого методу кодування є оперативність та точності обробки, а також та ресурсоємість кодеку;
- зростання рівня точності обробки є необхідним для створення умов, за яких у процесі кодування відеоінформації забезпечується достатня величина коефіцієнту стиснення, одночасно з чим якість реконструйованих даних суттєво не зменшується;
- відеоінформаційний потік, має пульсуючий закон змін у часі, що зумовлюється, з одного боку, особливостями змісту окремих кадрів а з іншого боку – закономірністю процесу обробки у схемі MPEG, що сьогодні є домінуючою;
- виходячи з особливостей відео, як специфічного потоку даних, для його кодування більш прийнятним є VBR-режим кодування, у той час як

у режимі CBR за використання заповнюючи пакетів має місце нераціональне використання пропускної здатності мережі;

- найбільш семантично значущими у межах кадру є контурні ділянки;
- максимальна концентрація пікселів у відновлених кадрах належить контурним ділянкам, обробка яких виконується з такими ж параметрами, як і інших ділянок у кадрі;
- підвищення ефективності обробки відео потоку за рахунок збільшення показника точності може бути реалізовано за рахунок окремого кодування фрагментів кадрів, що містять контурні ділянки.

3 РОЗРОБКА ПІДХОДІВ ДО КОДУВАННЯ НАЙБІЛЬШ СЕМАНТИЧНО ЗНАЧИМИХ ДІЛЯНОК ВІДЕОІНФОРМАЦІЙНОГО ПОТОКУ

3.1 Алгоритми пошуку контурних ділянок у кадрі

Будь-який алгоритм, що застосовується для виявлення ділянок того чи іншого типу у межах відеокадру, як двовимірного масиву даних, має відповідати таким вимогам, як [11]:

- однозначність розпізнавання контуру у межах тієї чи іншої локальної ділянки;
- урахування специфіки формування фрагментів відеокадрів;
- рівень складності реалізації, що дозволяє ефективно застосовувати механізм на базі будь-якої апаратної платформи.

У загальному випадку зараз існує 2 групи алгоритмів пошуку контурів у межах зображень, а саме:

- загальні;
- спеціалізовані.

До першої групи алгоритмів відносяться такі, що базуються на виявленні локальних ділянок високого рівня неоднорідності у межах оброблюваного відеозображення.

У свою чергу, другу групу складають алгоритми, що орієнтовані на обробку зображення як єдиного цілого, без передуючого опису множиною менших за розмірами фрагментів.

Сюди входять:

- лінійні алгоритми;
- нелінійні алгоритми.

Прикладами лінійних алгоритмів є маски, подібні тим, що застосовуються для лінійної фільтрації.

Відмінність з ними полягає у тому, що виявлення меж здійснюється з використанням вагових коефіцієнтів, що мають різні знаки.

Прикладами масок, що застосовуються для виконання горизонтальної та вертикальної обробки (підсилення, акцентування) зображення, можуть бути наступні:

$$\begin{Vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{Vmatrix} \text{ та } \begin{Vmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{Vmatrix}$$

Так, у межах однорідних зон кадру діапазон яскравості пікселів є невеликим, тобто, величина градієнту яскравості низька.

Таким чином, після виконання обробки такі ділянки кадру будуть лише мати нижчий рівень яскравості, тобто, будуть візуально темнішими.

Разом з тим, на ділянках різких перепадів крутизна зміни рівня яскравості є досить високою. Це, у свою чергу, веде до появи яскравих ділянок на підсумкових зображеннях.

Означені алгоритми не набули широкої практичної реалізації як наслідок того, що частина з них демонструє високу ефективність лише у певних обмежених умовах (зумовлених діапазоном змін яскравості у межах зображень, що обробляються).

Натомість, нелінійні алгоритми сьогодні використовуються для обробки статичних зображень, що дозволяє застосовувати їх також для виявлення контурів у межах динамічної структури – потоку відеокадрів.

Серед нелінійних алгоритмів найчастіше використовуються операнди Собеля та Робертса, операнди Превіта та подібні.

Розглянемо більш детально сутність алгоритмів загальної групи та нелінійних алгоритмів з групи спеціалізованих.

3.2 Сутність алгоритмів загальної групи виявлення контурів

Контуром вважається умовна межа переходу між двома середовищами [18, 19].

Така межа має задовольняти вимоги щодо різниці λ однієї з характеристик сусідніх середовищ, а саме – повинна бути не меншою, ніж деякий пороговий рівень λ_{th}, тобто $\lambda \geq \lambda_{th}$.

Стосовно зображень – як статичних, так і динамічних їх масивів – такою характеристикою є яскравість δ ділянок, що знаходяться безпосередньо одна біля одної.

Таким чином, ймовірною ознакою локалізації контуру у межах фрагменту відеопотоку при його розгляді у просторовій площині може бути певний рівень різниці величин яскравості $\delta_{i,j}$ пікселів.

Далі необхідно визначити, яку саме структурну одиницю відео інформаційного потоку найбільш доцільно використати у якості ключової, тобто, у рамках якої далі буде виконуватися обробка.

3.2.1 Декомпозиція відеопотоку до різних рівнів представлення для виявлення найбільш ефективних для виявлення контурної інформації

Найвищим рівнем декомпозиції відеоінформаційного потоку є група кадрів (GOP). У нашому ж випадку мова йде про обробку окремих кадрів у потоці.

У свою чергу, вибір кадру F, як ключової одиниці відеоінформаційного потоку, на рівні якого виконується пошук контурів, є також недоцільним, так як за мови, що гіпотетично відома величина δ_{th} порогу, може бути виявлено лише факт наявності контурної інформації у межах кадру, чи її відсутність.

З таких же міркувань недоцільно виконувати процедуру пошуку контурних даних на рівні декомпозиції кадру F до сукупності тайлів та слайсів.

Разом з тим, у якості базового, або першого рівня декомпозиції відеопотоку, починаючи з якого доцільно вести пошук контурів, є сукупність локальних фрагментів $f_{x,y}^{(F)}$ розміром 8x8 пікселів кожен, що є традиційним для JPEG.

При цьому, вихідний кадр F розглядається як деяка сукупністю локальних фрагментів $f_{x,y}^{(F)}$, тобто:

$$F = \bigcup_{x=1}^{X} \bigcup_{y=1}^{Y} f_{x,y}^{(F)}, \qquad (3.1)$$

де X та Y – кількість локальні фрагментів у рядку та стовпцю кадру відповідно.

3.2.2 Процедура виявлення контурів на рівні фрагментів кадру

На рівні фрагментів $f_{x,y}^{(F)}$ процедура виявлення контурів, що базується на відстежуванні різниць компонент яскравості, може бути описана наступною системою виразів [20]:

$$\begin{cases} \delta_1 = \sum_{k=1}^{\nu_c} \delta_{k,\ell}^{(max)} - \delta_{k,\ell}^{(min)} \\ \delta_2 = \sum_{\ell=1}^{\upsilon_c} \delta_{k,\ell}^{(max)} - \delta_{k,\ell}^{(min)} \end{cases}, \tag{3.2}$$

де $\delta_{k,\ell}^{(max)}$ та $\delta_{k,\ell}^{(min)}$ - найбільша та найменша яскравісна компоненти у межах а-го рядку відповідно.

Як видно з виразу 3.2, для кожного фрагменту $f_{x,y}^{(F)}$ першого рівня декомпозиції кадру F розраховується 2 показника суми різниць величин яскравісних компонент – за рядками, та за стовпцями.

Такий підхід виключає ймовірність помилкового розпізнавання фрагменту $f_{x,y}^{(F)}$ як такого, що не містить контурів.

При цьому, подібне некоректне розпізнавання можливе тоді, коли контур має чітку односпрямовану локалізацію – горизонтальну, або вертикальну.

Для такого випадку показник δ може бути суттєво нижчим, ніж попередньо встановлена величина δ_{th} порогу.

Отже, у підсумку кожному локальному фрагменту $f_{x,y}^{(F)}$ першого рівня декомпозиції кадру F ставиться у відповідність пара значень δ_1 та δ_2.

При цьому, якщо хоча б одне з одержаних значень - δ_1, або δ_2, буде не меншим, ніж рівень порогу δ_{th}, надалі вважається, що фрагмент $f_{x,y}^{(F)}$ містить контурні дані, інакше кажучи – належить множині C контурних, тобто:

$$\begin{cases} f_{x,y}^{(F)} \in C \mid \delta_1 \vee \delta_2 \geq \delta_{th} \\ f_{x,y}^{(F)} \in C \mid \delta_1 \wedge \delta_2 \geq \delta_{th} \\ f_{x,y}^{(F)} \notin C \mid \delta_1 \wedge \delta_2 < \delta_{th} \end{cases}. \quad (3.3)$$

Після виконання даної операції відносно відеокадру, далі він розглядається як конкатенація множини C фрагментів контурного типу, та множини H фрагментів, що не містять контурів:

$$F = C \& H. \quad (3.4)$$

Такий поділ локальних фрагментів $f_{x,y}^{(F)}$ кадру на множини має сенс для подальшого контролю рівня K_{enc} коефіцієнта стиснення.

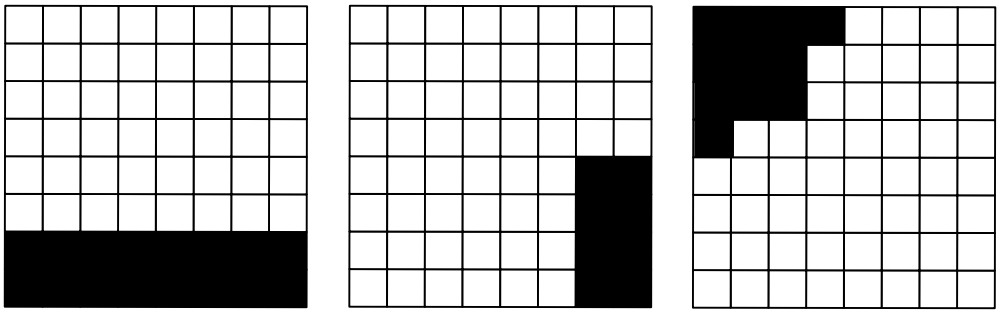
У загальному випадку, передбачається, що під час подальшої обробки його величина встановлюється залежно від того, який саме поточний локальний фрагмент $f_{x,y}^{(F)}$ підлягає кодуванню.

Даний принцип може бути представлено наступним виразом:

$$\begin{cases} K_{enc} := K_{enc} - \gamma_1 \mid f_{x,y}^{(F)} \in C \\ K_{enc} := K_{enc} + \gamma_2 \mid f_{x,y}^{(F)} \in H \end{cases}, \quad (3.5)$$

де γ_1 та γ_2 - величини, на яку змінюється значення K_{enc}, що задане за замовчуванням для базових умов обробки.

У свою чергу, локалізація контуру у межах фрагментів $f_{x,y}^{(F)}$ може бути такою, коли більша частина фрагменту представлена даними текстурного чи градієнтного характеру, або однотонною чи квазі-однотонною ділянкою, тоді як контурною інформацією фрагмент буде заповнюватися незначно (рис.3.1).



а) б) в)

Рисунок 3.1 – Приклади локалізації контурів у межах базових фрагментів кадру

Особливістю наведених рис.3.1 часткових прикладів позиціонування контурної інформації у фрагментах $f_{x,y}^{(F)}$ є те, що у ряді випадків за результатами обробки на базі виразу (3.2) можуть виникнути умови, коли:

$$\begin{cases} \delta_1 \wedge \delta_2 \approx \delta_{th} \;\&\; \delta_1 \wedge \delta_2 < \delta_{th} \\ \delta_1 \vee \delta_2 \approx \delta_{th} \;\&\; \delta_1 \vee \delta_2 < \delta_{th} \end{cases}. \tag{3.6}$$

Тобто, виникає ситуація невизначеності, коли фрагмент $f_{x,y}^{(F)}$, який відповідає умовам (3.6), позначається як квазі-контурний, тобто, таких, що може містити контурні дані. Інакше кажучи, висувається гіпотеза про можливість локалізації контурів у досліджуваному фрагменті.

Для вирішення такої невизначеності щодо належності фрагменту $f_{x,y}^{(F)}$ одній з множин – C чи H, далі відносно нього виконується процедура додаткової декомпозиції.

За результатами цього утворюється сукупність субфрагментів, кожен з яких має розміри $\frac{X}{2} \times \frac{Y}{2}$, що у випадку базових розмірів 8×8 фрагменту $f_{x,y}^{(F)}$ рівняється 4×4, як показано рис.3.2.

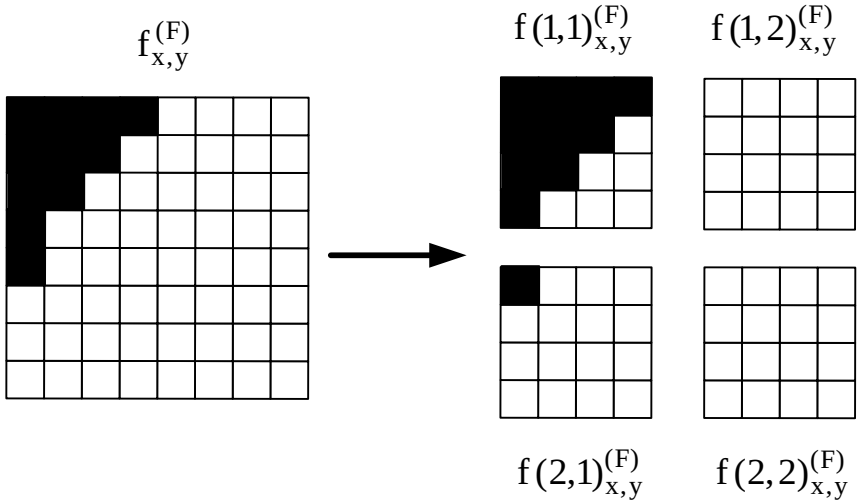


Рисунок 3.2 – Приклад утворення масиву субфрагментів за рахунок декомпозиції базового фрагменту $f_{x,y}^{(F)}$ кадру F

Далі кожен з субфрагментів - $f(1,1)_{x,y}^{(F)}$, $f(1,2)_{x,y}^{(F)}$, $f(2,1)_{x,y}^{(F)}$ та $f(2,2)_{x,y}^{(F)}$ підлягає повторному аналізу на базі виразу (3.2).

За результатами такого аналізу гіпотезу про можливість існування контурів усередині фрагменту $f_{x,y}^{(F)}$ буде або підтверджено, або спростовано. Після цього субфрагменти, для яких виконуються умови (3.2) маркуються, як такі, що належать множині контурних. Так, на рис.3.2 субфрагмент $f(1,1)_{x,y}^{(F)}$ визнається контурним, тоді як $f(1,2)_{x,y}^{(F)}$, $f(2,1)_{x,y}^{(F)}$ та $f(2,2)_{x,y}^{(F)}$ - такими, що контурної інформації не містять.

Важливо зазначити, що величина порогового значення для субфрагментів $f(\alpha,\beta)_{x,y}^{(F)}$ приймається рівним $\frac{1}{2}\delta_{th}$.

За результатами виконаних дій до множини C вноситься підмножина C' контурних субфрагментів, отриманих на другому рівні декомпозиції кадру F. Тоді вираз (3.4) буде уточнено до вигляду:

$$F = \{C \ \& \ C'\} \ \& \ H. \qquad (3.7)$$

Надалі обробка множин H, C та C' виконується згідно принципу, наведеному виразом (3.5). Тобто, для множини H застосовується збільшений коефіцієнт K_{enc} стиснення, а для множини C та підмножини C' - зменшений.

Для випадку застосування розглянутого підходу до виявлення контурів на рівні першого та другого рівнів декомпозиції відеокадру F загальна схема буде такою, як зображено рис.3.3.

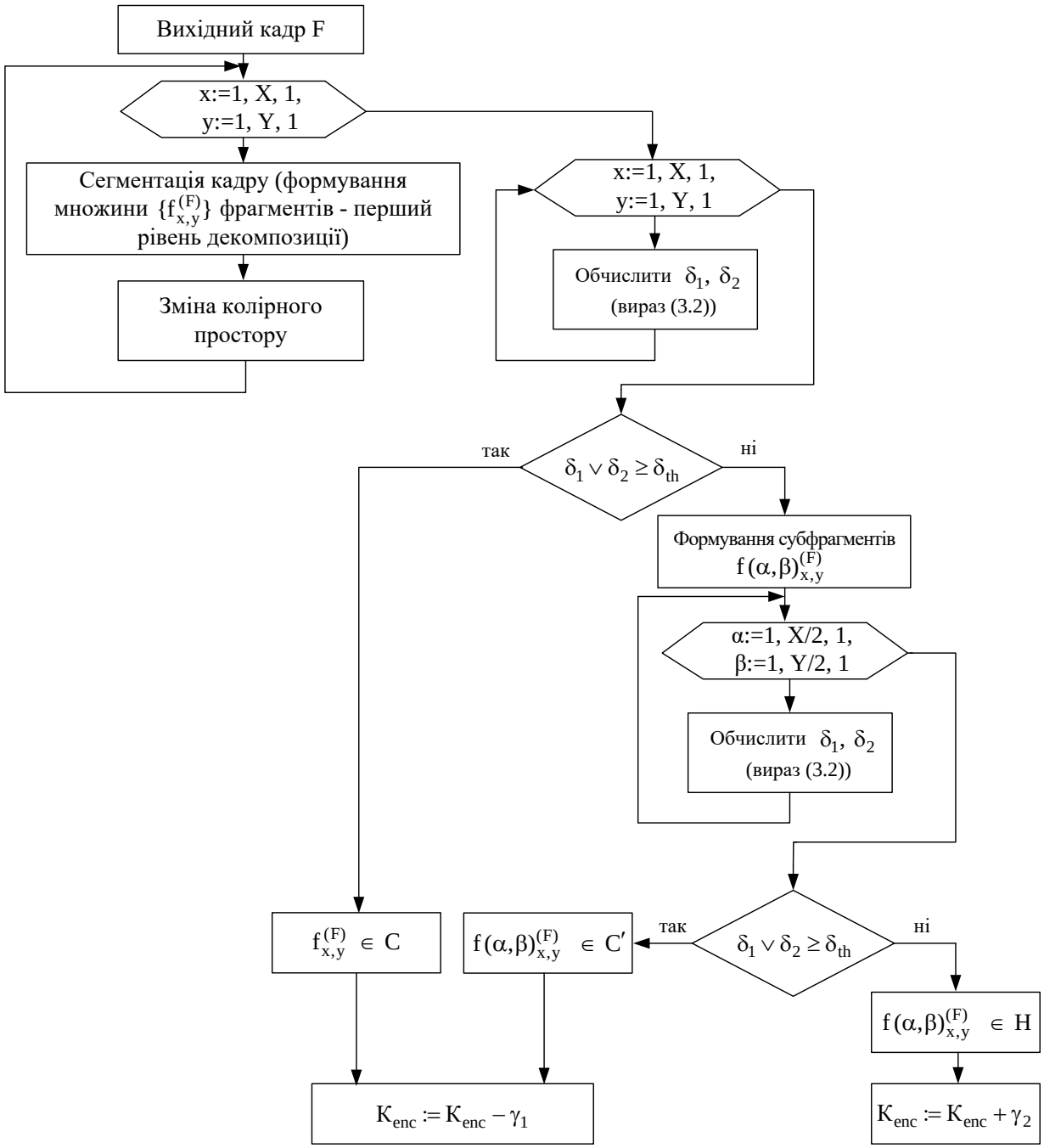


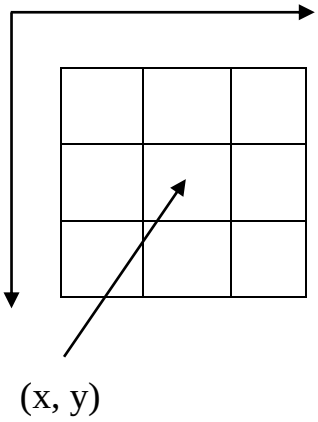
Рисунок 3.3 – Загальна схема виявлення контурів за участю першого та другого рівнів декомпозиції кадру

3.3 Виявлення контурів у зображеннях на базі алгоритмів спеціалізованої групи

У рамках спеціалізованих алгоритмів також виділення контурів здійснюється за величиною рівня компонент $\delta_{k\ell}$ яскравості, керуючись однієї з ключових властивостей яскравісного сигналу – властивості розривності [19, 21].

При цьому, найбільш широко застосовуваний метод пошуку ділянок зміни яскравості полягає у обробці компонент яскравості $\delta_{x,y}$ з використанням ковзаючої маски (рис.3.4), що являє собою матрицю розміром nxn, та також має назву ядра фільтрації, фільтру, вікна, шаблону чи операнду.

У свою чергу, процес маніпуляції відносно зображення з використанням таких матриць може зватися також фільтрацією.



Маска фільтру

z_1 $\delta_{x-1,y-1}$	z_2 $\delta_{x-1,y}$	z_3 $\delta_{x-1,y+1}$
z_4 $\delta_{x,y-1}$	z_5 $\delta_{x,y}$	z_6 $\delta_{x,y+1}$
z_7 $\delta_{x+1,y-1}$	z_8 $\delta_{x+1,y}$	z_9 $\delta_{x+1,y+1}$

Коефіцієнти маски

Рисунок 3.4 – Ковзаюча маска-фільтр (операнд), що використовується для пошуку ділянок зміни яскравості

Для випадку застосування методів фільтрації для виявлення контурної інформації у зображенні поняття контуру має бути уточнено. Його сутність може бути проілюстрована рис.3.5.

При цьому, у будь якому випадку контур також початково розглядається як певна кількість пікселів, між якими значення компонент $\delta_{x,y}$ яскравості суттєвим чином змінюються.

Разом з тим, тут характер зміни величини компонент яскравості може бути описано такими показниками, як:

- величина переходу $\delta_{0;}$
- кут нахилу θ;
- координата δ_0 центру похилої ділянки.

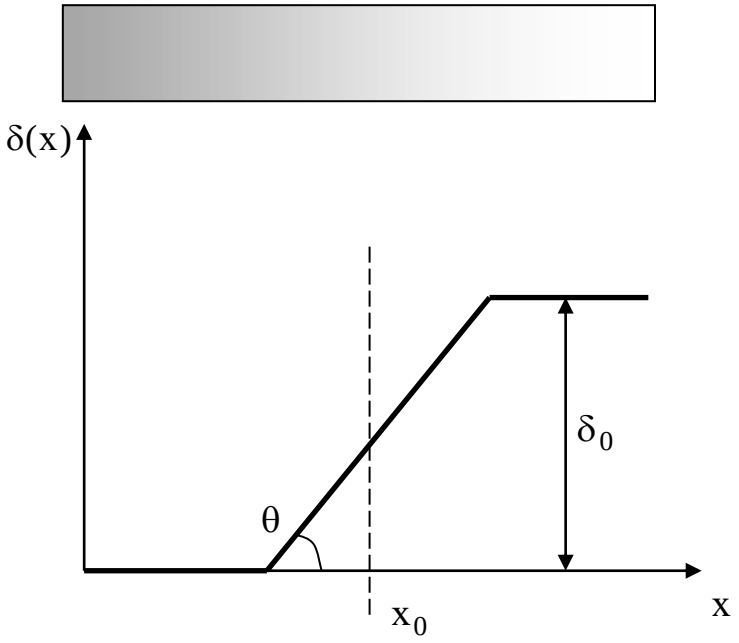


Рисунок 3.5 – Схематична інтерпретація поняття контуру

Існуючий перехід між значеннями $\delta_{x,y}$ яскравості може вважатися контуром за тієї умови, якщо величина δ_0 разом з кутом нахилу θ є більшими, ніж деякі встановлені порогові значення $\delta_0^{(th)}$ та $\theta^{(th)}$, тобто:

$$\begin{cases} \delta_0 \geq f_0^{(th)} \\ \theta \geq \theta_{th} \end{cases}. \qquad (3.8)$$

У загальному вигляді, аналітична форма опису інтенсивності зміни рівнів яскравості за напрямками x_1 та x_2 передбачає використання похідних $\frac{df}{dx_1}$ та $\frac{df}{dx_2}$.

Обчислення таких похідних у довільних напрямках може здійснюватися на базі величин градієнтів яскравості (рис.3.6).

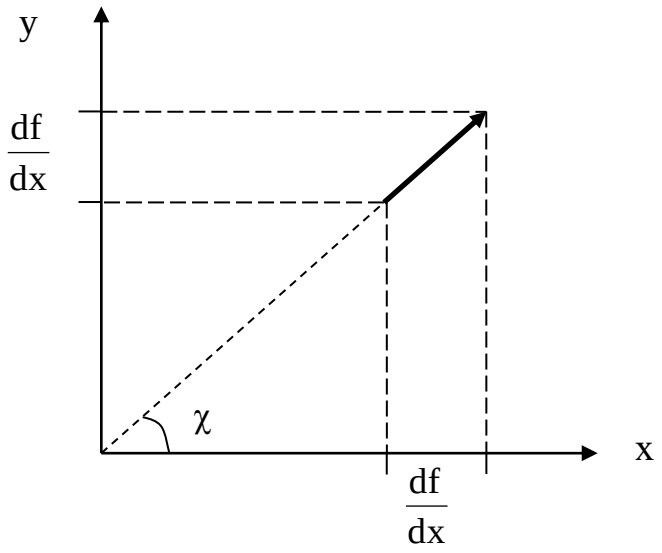


Рисунок 3.6 – Графічний опис градієнту яскравості

У ході фільтрації, спрямованої на виявлення контурів, здійснюється поступове зміщення операнду у межах двомірного масиву компонент $\delta_{x,y}$ яскравості з одночасним знаходженням на кожній позиції (x,y) величини відгуку Θ фільтру. Дана операція обчислення потребує використання попередньо заданих зв'язків. Так, на випадок лінійної фільтрації (що застосовується у 90% випадків), відгук обчислюється, як сума добутків елементів фильтру на величини компонент $\delta_{x,y}$ яскравості, що відповідають їм у зонах, які перекриваються операндом.

Розглянемо для прикладу операнду розміром 3x3 елемента, як показано рис.3.4. Для цього випадку величина відгуку Θ фільтрації, отриманого у точці (x,y), може бути отримано на базі наступного виразу:

$$\begin{gathered}\Theta = z_1\delta_{x-1,\,y-1} + z_2\delta_{x-1,\,y} + z_3\delta_{x-1,\,y+1} + \\ + z_4\delta_{x,\,y-1} + z_5\delta_{x,\,y} + z_6\delta_{x,\,y+1} + z_7\delta_{x+1,\,y-1} + \\ + z_8\delta_{x+1,\,y} + z_9\delta_{x+1,\,y+1}\end{gathered} \qquad (3.9)$$

Тобто, як видно з виразу (3.9), відгук Θ являє собою суму добутків елементів операнду на величини компонент $\delta_{x,y}$ яскравості безпосередньо під ним. Важливим тут є те, що елементу z_5, що є центральним у операнді, відповідає компонента $\delta_{x,y}$, тим самим вказуючи на те, що операнд центрується відносно точки (x,y).

Для виявлення різких перепадів рівня яскравості при цьому використовуються дискретні аналоги похідних.

Так, для обчислення першої похідної одномірної функції δ_x, виконується операція розрахунку різниці сусідніх компонент, а саме:

$$\frac{df}{dx} = \delta_{x+1} - \delta_x . \tag{3.10}$$

При цьому, тут задіяно вираз у вигляді часткової похідної.

Даний підхід є важливим для подальшого збереження таких же позначень на випадок застосування двох змінних $\delta_{x,y}$, коли необхідно розраховувати часткові похідні за двома координатними осями.

Разом з тим, застосування розрахунків на базі часткових похідних не змінює сутності процесу.

У свою чергу, за аналогією з виразом (3.10), розрахунок другої похідної зводиться до знаходження різниці сусідніх значень першої похідної, тобто:

$$\frac{d^2f}{dx^2} = \delta_{x+1} - \delta_{x-1} + 2\delta_x . \tag{3.11}$$

Тут розрахунок першої похідної потребує виконання ряду дискретних наближень градієнту у двовимірному просторі.

При цьому береться до уваги те, що за визначенням, - у межах зображення градієнт величини $\delta_{x,y}$ на позиції (x,y) розглядається як вектор, аналітичний вираз для якого може бути зображено виразом:

$$\nabla\delta = \left[\frac{\Theta_x}{\Theta_y}\right] = \left[\frac{\frac{df}{dx}}{\frac{df}{dy}}\right]. \tag{3.12}$$

Як відомо, просторова спрямованість вектору градієнту та напрямку найбільшої швидкості змін значення функції δ у координаті (x,y) повністю збігаються [19].

У свою чергу, для модуль вектору $\nabla\delta$ є необхідним для виявлення контурів.

Для його розрахунку використовується наступний вираз:

$$|\nabla\delta| = \sqrt{\Theta_x^2 + \Theta_y^2} \tag{3.13}$$

Значення модулю вектору $\nabla\delta$ відповідає найбільшій швидкості зростання функції δ у точці з координатами (x,y). Також величина $\nabla\delta$ зветься вектором градієнту.

Що одним ключовим параметром, що використовується у ході виявлення контурів, є напрямок вектору градієнта. Для його визначення необхідно обчислити кут χ (рис.3.6), що утворюється між напрямком вектору $\nabla\delta$ у точці, яка має координату (x,y), а також віссю x. Величину даного кута може бути обчислено на базі формули:

$$\chi = \operatorname{arctg}\left(\frac{\Theta_y}{\Theta_x}\right) \tag{3.14}$$

Спрямованість контуру у точці з координатою (x,y) є перпендикуляром відносно напрямку вектору $\nabla\delta$ на даній просторовій позиції. У свою чергу, градієнт яскравості обчислюється на базі значень часткових похідних $\frac{df}{dx}$ а також $\frac{df}{dy}$ у кожній точці.

Далі розглянемо найбільш поширені маски просторової фільтрації, що застосовуються у процесі пошуку контурів у зображеннях. Одним з найбільш простих та часто застосовуваних масок є перехресний градієнтний операнд Робертса [19, 21].

3.3.1 Операнд Робертса

Операнд Робертса у ході просторової фільтрації використовує 4 сусідні відносно точки (x,y) пікселі, які щодо даної точки локалізовані так, як показано рис.3.7 [18].

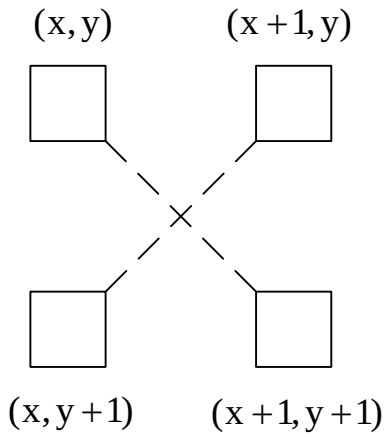


Рисунок 3.7 – Схема розрахунку перших часткових похідних у точці (x, y) на базі використання перехресного градієнтного оператору Робертса

Розрахунок перших часткових похідних відносно пікселя z_5 означеним способом при цьому виконується за формулами [18]:

$$\Theta_x = (z_9 - z_5), \tag{3.15}$$

а також

$$\Theta_x = (z_8 - z_6). \tag{3.16}$$

Власне, самі операнди Робертса є такими, як зображено рис.3.8.

-1	0
0	1

0	-1
1	0

Рисунок 3.8 – Операнди Робертса, що використовуються для виявлення контурів

Розмірність масок 2x2 дозволяє виконувати обробку з мінімальними часовими затримками та мінімальним обчислювальним навантаженням на апартну платформу. У той же час, такий підхід має недолік, що пояснюється

відсутністю центра маски, що, у свою чергу, сприяє частковому зменшенню якості обробки.

Аналогічно використанню операндів Робертса, може бути застосовано операдни Превіта та Собеля.

3.3.2 Операнд Собеля

Ключовою відмінністю як операнду Собеля, так і Превіта є його розмірність 3х3 (рис.3.9), що, у свою чергу, дозволяє виконувати контурний аналіз зображення на рівні локальних його зон такого ж розміру.

Окрім розмірності самої маски, ще одна відмінність операнду Собеля щодо операнду Робертса полягає у використанні вагових коефіцієнтів рівних 2 для середніх елементів шаблону, як показано рис. 3.9.

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Рисунок 3.9 - Операнди Собеля, що використовуються у процесі пошуку контурів

Відповідно, для даного операнду розрахунок відгуків фільру здійснюється на базі наступних виразів:

$$\Theta_x = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3), \tag{3.17}$$

а також

$$\Theta_y = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7). \tag{3.18}$$

Сенс використання збільшених величин вагових коефіцієнтів полягає у тому, що за рахунок цього досягається помітне зменшення ефекту розмиття

контуру. Це викликано тим, що середнім точкам передається більший рівень ваги, ніж у випадках інших операндів. З цієї точки зору виявлення контурів з використанням оператору Собеля у ряді випадків є ефективнішим, ніж схожі шаблони.

Отже, головна функція операндів у процесі виявлення контурів полягає у знаходженні відгуків Θ_x та Θ_y фільтрації. Далі на їх базі розраховуються ключові складові градієнту відповідно до виразів (3.13) та (3.14).

3.3.3 Операнд Превіта

Даний операнд, як і у випадку операнда Собеля, має розмірність 3x3. Проте сам операнд відрізняється від попереднього, та має вигляд, як показано рис.3.10.

-1	-1	-1
0	0	0
1	1	1

-1	0	1
-1	0	1
-1	0	1

Рисунок 3.10 - Операнди Превіта

Обчислення відгуків фільтру Θ_x та Θ_y з використанням даного операнду виконується наступним чином:

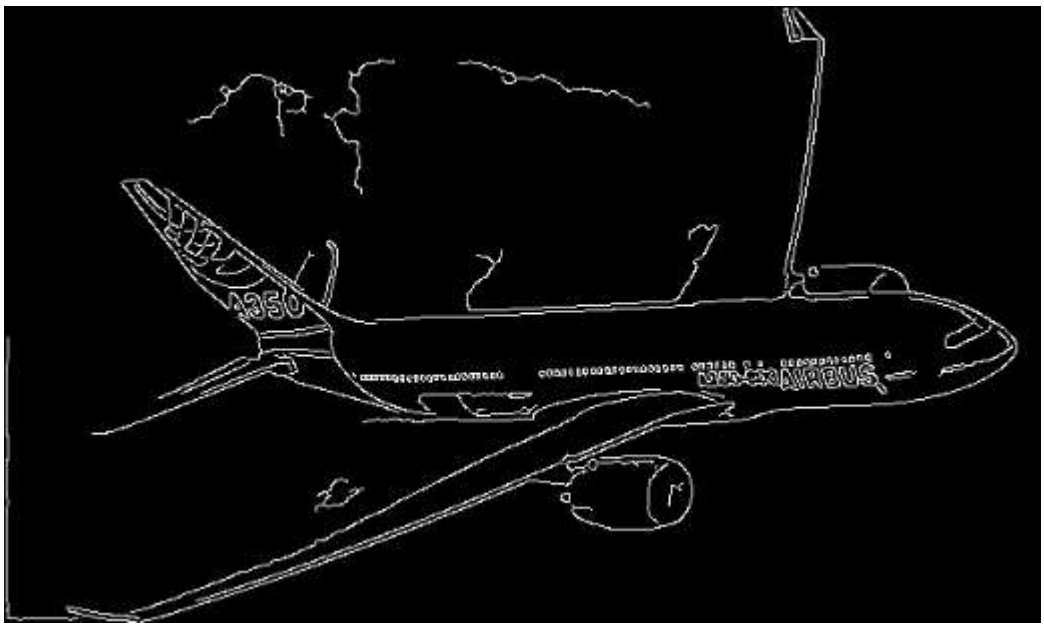
$$\Theta_x = (z_7 + z_8 + z_9) - (z_1 + z_2 + z_3), \tag{3.19}$$

та

$$\Theta_y = (z_3 + z_6 + z_9) - (z_1 + z_4 + z_7). \tag{3.20}$$

Серед розглянутих шаблонів, призначених для пошуку контурів у зображеннях, найбільш точні результати забезпечує операндів Собеля. Стосовно розглянутих на рис.2.5 тестових кадрів, результати виявлення контурів на його базі демонструються рис.3.11. Для виявлення контурної інформації з використанням операндів Собеля застосовано програмний засіб «Калькулятори» онлайн-сервісу Upbyte [22].

а)



б)

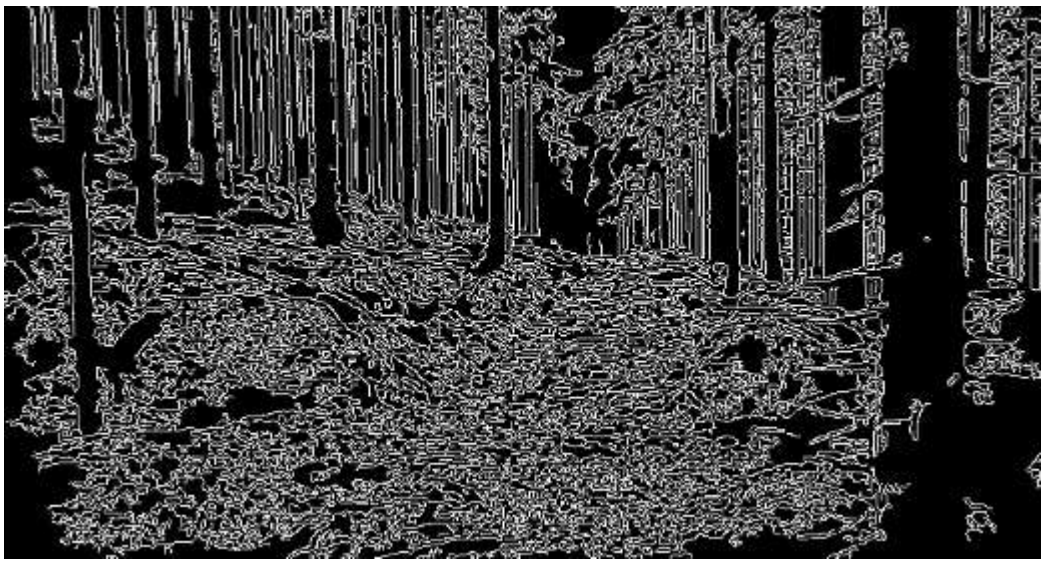


Рисунок 3.11 – Результати виявлення контурних даних на тестових кадрах – Кадр 1 а) Кадр 2 б)

Таким чином, рис.3.11 наочно демонструє відмінність у структурі тестових кадрів, пояснює відмінність результатів їх кодування зі сталим коефіцієнтом стиснення та доводить доцільність окремої обробки контурних ділянок та ділянок кадру, що контурів не містять у спосіб, що поданий виразом (3.5).

Далі виконаємо порівняльний наліз методів виявлення контурів загальної групи та спеціалізованих.

4 ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ КОНТУРІВ ЗАГАЛЬНОЇ ТА СПЕЦІАЛІЗОВАНОЇ ГРУП

4.1 Критерії порівняння методів виявлення контурів

Для визначення найбільш доцільного підходу для виявлення контурів під час кодування відео у реальному часі, порівняємо розглянуті групи методів за такими критеріями порівняння, як:

- швидкість обробки (обсяг необхідних арифметичних операцій);
- достовірність отриманих результатів;
- функціональні особливості груп методів.

4.1.1 Розрахунок кількості необхідних арифметичних операцій для обробки фрагменту зображення

У загальному випадку, безвідносно потужності обчислювальної платформи, швидкість ν_{perf} обробки структурної одиниці відеоінформаційного потоку є зворотно пропорційною кількості E арифметичних операцій, які мають при цьому виконуватися, тобто [23]:

$$\nu_{perf} = \frac{1}{E} \qquad (4.1)$$

Методи першої групи орієнтовані на виявлення самого факту присутності даних контурного типу без детального розпізнавання контуру. При цьому, для їх випадку обробка виконується на рівні фрагментів $f_{x,y}^{(F)}$ першого чи другого $f(\alpha,\beta)_{x,y}^{(F)}$ рівня декомпозиції кадру прямим зчитуванням значень $\delta_{x,y}$ компонент яскравості з наступним обчисленням за виразом (3.2) величин δ_1 та δ_2.

При цьому, якщо не брати до уваги операцій запису у чарунки пам’яті, запису та присвоєння, що є найменш ресурсоємними і відтак не складають помітного впливу на складність обчислень, за найгірших умов на першому рівні декомпозиції кадру, як свідчить вираз (3.2) для обробки фрагменту $f_{x,y}^{(F)}$ виконується:

- $E^{(or)}$ операцій «логічне АБО» (у ході пошуку $\delta_{x,y}^{(max)}$ та $\delta_{x,y}^{(min)}$);

- $E^{(add)}$ операцій «add», або додавання (для обчислення суми різниць значень компонент);

- $E^{(sub)}$ операцій «sub», або відмімання.

При цьому, їх загальна кількість E для будь якого рівня декомпозиції кадру буде дорівнювати:

$$E = \left(E^{(or)} + E^{(add)} + E^{(sub)}\right) \times 2, \tag{4.2}$$

Множник, рівний 2, пояснюється тим, що обробка виконується за 2 проходи – за рядками та за стовпцями, як показано виразом (3.2).

У свою чергу, згідно з (3.21) на першому рівні декомпозиції кадру виконується така кількість операцій «логічне АБО»:

$$E^{(or)} = \frac{n}{\rho} \times \frac{n}{\rho} + \frac{n/\rho - 1}{n/\rho} \times \frac{n}{\rho} \times \frac{n}{\rho} + 1, \tag{4.3}$$

де ρ - індекс рівня декомпозиції кадру;

$\frac{n/\rho - 1}{n/\rho} \times \frac{n}{\rho} \times \frac{n}{\rho}$ - складник, що показує кількість операцій «логічне АБО» на другому проході компонент фрагменту $f_{x,y}^{(F)}$ під час пошуку їх максимальних (мінімальних) значень. Сутність множника $\frac{n/\rho - 1}{n/\rho}$ полягає у тому, що у ході першого проходу максимальні (мінімальні) величини компонент вже виявлено, відтак на другому проході 1 компонента з n не обробляється.

У свою чергу, доданок 1 тут пояснюється тим, що одержана величина δ_1 (та δ_2 у ході другого проходу) порівнюється з δ_{th}, на що витрачається одна операція «логічне АБО».

Тоді $E^{(or)} = 8 \times 8 + \frac{8-1}{8}(8 \times 8) = 120$ операцій «логічне АБО».

Відповідно, для фрагменту nxn витрачається така кількість операцій додавання, як:

$$E^{(add)} = \frac{n}{\rho} - 1, \tag{4.4}$$

Тобто, на рівні фрагменту $f_{x,y}^{(F)}$ маємо $E^{(add)} = 7$.

Разом з тим, кількість операцій $E^{(sub)}$ з виразу (3.2) визначатиметься як:

$$E^{(sub)} = \frac{n}{\rho} \tag{4.5}$$

Отже, з виразу (4.5) $E^{(sub)} = 8$. Таким чином, $E = (120 + 7 + 8) \times 2 = 270$ арифметичних операцій.

На найгірший випадок, тобто, тоді, коли на першому рівні декомпозиції присутність контуру у досліджуваному фрагменті $f_{x,y}^{(F)}$ не доведено чітко (виконуються умови (3.6)), додатковий необхідний обсяг E_2 арифметичних операцій на другому рівні декомпозиції кадру може бути оцінено, аналогічно (3.21), за таким виразом:

$$E_2 = \left(E_2^{(or)} + E_2^{(add)} + E_2^{(sub)} \right) \times 8. \tag{4.6}$$

Присутність множника 8 у виразі (4.6) зумовлюється, по-перше, тим, що складові виразу у дужках дозволяють розрахувати сукупність арифметичних дій для одного субфрагменту $f(\alpha,\beta)_{x,y}^{(F)}$ другого рівня декомпозиції, яких у межах фрагменту $f_{x,y}^{(F)}$ міститься 4. При цьому, кожен з них обробляється на базі виразу (3.2) 2 рази – у напрямку рядків та у напрямку стовпців.

При цьому, у відповідності з (4.5), $E_2^{(sub)} = \frac{8}{2} = 4$, а $E_2^{(add)} = \frac{8}{2} - 1 = 3$.

У свою чергу, $E_2^{(or)}$ для другого рівня декомпозиції дорівнює $E_2^{(or)} = \frac{8}{2} \times \frac{8}{2} + \frac{8/2 - 1}{8/2} \times \frac{8}{2} \times \frac{8}{2} + 1 = 29$. Відтак, загальна кількість

арифметичних операцій, необхідних для виявлення контурної інформації на другому рівні декомпозиції дорівнює $E_2 = (29+4+3)\times 8 = 288$.

У підсумку, за умови, що фрагмент $f_{x,y}^{(F)}$ кадру потребує виконання процедури виявлення наявності контурів на двох рівнях декомпозиції, загальний обсяг арифметичних операцій , який витратиться для цього, є рівним:

$$E_\Sigma = E + E_2. \qquad (4.7)$$

Отже, тоді $E_\Sigma = 270 + 288 = 558$.

Величина E_Σ є апріорі фіксованою та фактично не залежить від особливостей змісту фрагментів $f_{x,y}^{(F)}$ кадру. Виняток складають умови (3.6), коли пошук контурних даних необхідно виконувати на другому рівні декомпозиції.

Виконаємо далі обчислення необхідного обсягу арифметичних операцій на випадок застосування методу пошуку контурів з використанням ковзаючої маски на базі розглянутого операнду Собеля для найгіршого випадку.

Найгіршим випадком можуть вважатися умови, коли контур має складну структуру, а у межах основного контуру може бути локалізовано зовнішні контури, що не перекриваються з поточним (рис.4.1). У таких умовах може потребуватися обробка усіх наявних пікселів [21].

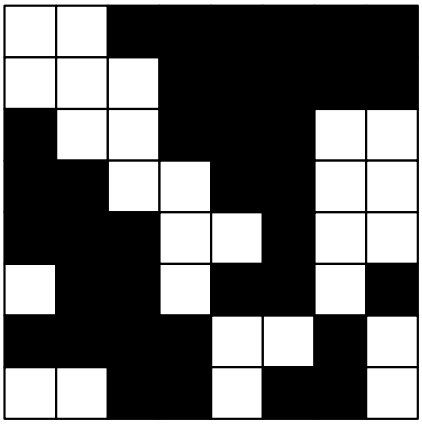


Рисунок 4.1 – Приклад складних примикаючи контурів

Так, за аналогією з методами, спрямованими на виявлення факту присутності контуру у межах фрагменту $f_{x,y}^{(F)}$, процес виявлення контуру з використанням операнду Собеля міститиме операції:

- додавання (add) $E_{sob}^{(add)}$;
- віднімання (sub) $E_{sob}^{(sub)}$;
- логічне АБО (or) $E_{sob}^{(or)}$;
- ділення (div) $E_{sob}^{(div)}$;
- множення (mul) $E_{sob}^{(mul)}$;
- арктангенс $E_{sob}^{(atg)}$;
- квадратний корінь $E_{sob}^{(sqr)}$.

Якщо вести розрахунок з точки зору групування означених арифметичних дій за певними технологічними етапами розрахунку, загальна кількість операцій буде обчислюватися як :

$$E_{sob} = E_{sob}^{(fil)} + E_{sob}^{(cnr)} + E_{sob}^{(mod)} + E_{sob}^{(or)}, \tag{4.8}$$

де $E_{sob}^{(fil)}$ - обсяг операцій для обчислення відгуків фільтра;

$E_{sob}^{(cnr)}$ - кількість дій для визначення кута градієнта;

$E_{sob}^{(mod)}$ - кількість дій для розрахунку модуля вектору градієнта;

$E_{sob}^{(or)}$ - кількість дій, необхідних для порівняння отриманого значення з пороговим.

При цьому, процес обчислення відгуків Θ_x та Θ_y фільтру за виразами (3.17) та (3.18) витрачає:

- 4 операції додавання;
- 1 операцію віднімання;
- 2 операції множення.

Таким чином, $E_{sob}^{(fil)} = 7$ для випадку одного пікселя. Відповідно для фрагменту $f_{x,y}^{(F)}$ маємо $E_{sob}^{(fil)} = 7 \times 64 = 448$

У свою чергу, у процесі розрахунку модулю вектора градієнта за виразом (3.13) виконується:

- 1 операція додавання;
- 2 операції множення.
- 1 операція обчислення квадратного кореня.

Оскільки, за умовами обчислень, підкореневий вираз у формулі (3.13) може приймати довільні значення, знаходження кореню за допомогою табличного підходу у даному випадку не може бути використано. При цьому, розрахунок квадратного кореню можливо виконувати або арифметично, або розкладанням числа у ряд Тейлора [18].

На випадок, коли необхідна точність обчислення модулю градієнту дозволяє виконувати розрахунки квадратного кореню наближено, таке обчислення може бути здійснено на базі однієї з характерних властивостей квадратів чисел, як показано далі:

$$\sum_{k=1}^{m}(2k-1)=m^2, \tag{4.9}$$

де m - корінь числа, який необхідно знайти.

Разом з тим, такий підхід має ряд недоліків, а саме:

- кількість операцій обчислень на базі виразу (4.9) зростає відповідно до збільшення числа m за законом арифметичної прогресії. Так, для m=3 маємо $3^2=1+3+5$. Відповідно, у процесі пошуку квадратного кореню виконується 3 операції віднімання ($9-1=8, 8-3=5$ і $5-5=0$). Таким же чином у випадку m=8 для розрахунку кореню виконується 8 операцій віднімання;
- метод дозволяє обчислити лише цілочисельну частину кореня та є неприйнятним для використання, коли за умовами обчислень необхідно забезпечити низький рівень похибки розрахунків.

Таким чином, на випадок можливості застосування арифметичного методу обчислення квадратного кореню у загальному випадку виконуватиметься $E_{sob}^{(mod)}=(1+2+1+m)\times 64=(m+4)\times 64$ арифметичних операції.

При цьому, загальна кількість $E_{sob}^{(cnr)}$ операцій для розрахунку кута χ градієнта за виразом 3.14 містить у собі:

- 1 операцію ділення (div) $E_{sob}^{(div)}$;

- 1 операцію визначення арктангенсу $E_{sob}^{(atg)}$.

Тут для розрахунку арктангенсу відношення Ξ відгуків фільтрації може бути використана формула наближеного обчислення:

$$\operatorname{arctg}(\Xi) \approx \Xi - \frac{\Xi^3}{3}. \tag{4.10}$$

Тоді маємо:

$$E_{sob}^{(atg)} = E_{sob}^{(sub)} + E_{sob}^{(div)} + E_{sob}^{(mul)} \tag{4.11}$$

Відповідно, для випадку, коли теоретично може оцінюватися кожна з точок фрагменту $f_{x,y}^{(F)}$ маємо $E_{sob}^{(atg)} = (1+1+3)\times 64 = 320$. Отже, $E_{sob}^{(cnr)} = (5+1)\times 64 = 384$

Таким чином, $E_{sob} = 448 + 384 + (m+4)\times 64 + 64 = 900 + (m+4)\times 64$.

Отже, навіть за умови m=1 кількість арифметичних операцій, що витрачаються у ході виявлення контуру у межах фрагменту $f_{x,y}^{(F)}$ майже двократно перевищує необхідну їх кількість для підходу, що орієнтований лише на виявлення факту наявності контурів.

4.1.2 Порівняння підходів до виявлення контурної інформації з позиції достовірності одержуваних даних

Нехай рівень достовірності у даному випадку буде оцінюватися за величиною D помилки, що вноситься у ході обробки та рівнем Ф точності обробки.

Як вже попередньо зазначалося, сутність порогових методів, що спрямовані на виявлення факту присутності контурів у межах тих чи інших структурних одиниць відео відеокадру, полягає у знаходженні перепадів яскравості δ вище деякого встановленого рівня δ_{th}.

Наявність таких перепадів, локалізованих на рівні деякого рівня декомпозиції структурних одиниць кадру може свідчити про існування контуру у їх межах.

При цьому, детальне розташування контуру та його структура є неважливими, що зумовлює відносно високу швидкість обробки фрагментів $f_{x,y}^{(F)}$ та її прогнозованість.

У той же час, достовірність розпізнавання присутності контурної інформації значною мірою залежить від встановленого рівня δ_{th} порогу.

На сьогодні можуть використовуватися такі підходи до встановлення порогу δ_{th}, як:

- фіксований;
- адаптивний (adatpive treshold).

При цьому, вибір того чи іншого підходу виконується з урахуванням існування пріоритетності або відносно точності Ф обробки, або відносно підсумкового рівня R інформаційної інтенсивності відеоінформаційного потоку.

Для першого підходу величина δ_{th} встановлюється автоматично системою. Це зазвичай деякий усереднений $\overline{\delta_{th}}$ рівень порогу, отриманий емпірично.

Зрозуміло, що керуючись величинами $\overline{\delta_{th}}$ ряд фрагментів $f_{x,y}^{(F)}$ та субфрагментів $f(\alpha,\beta)_{x,y}^{(F)}$ буде помилково розпізнано контурними чи такими, що кнурів не містять тоді, коли встановлений рівень $\overline{\delta_{th}}$ буде неприйнятним для даного типу зображення чи його частини (слайду або тайлу). У цьому випадку точність Ф обробки знизиться.

Так, оскільки ряд контурних фрагментів $f_{x,y}^{(F)}$ буде класифіковано як не маючих контурів, тоді, згідно виразу (3.5), їх кодування буде здійснюватися з використанням збільшеного коефіцієнту стиснення, тобто $K_{enc}+\gamma_2$.

Таким чином, інформаційна інтенсивність R результуючого потоку відео у середньому зменшиться. Очевидно також, що водночас зросте величина D внесеної помилки кодування.

Зазначена залежність є оберненою для випадку, коли деяка сукупність не містячи контурів фрагментів $f_{x,y}^{(F)}$ кодуватиметься з налаштуваннями, як для контурних. Це еквівалентно виразу:

$$\begin{cases} \overline{\delta_{th}} < \delta_{th}^{suit} \rightarrow D\uparrow, R\downarrow \rightarrow \Phi\downarrow \\ \overline{\delta_{th}} > \delta_{th}^{suit} \rightarrow D\downarrow, R\uparrow \rightarrow \Phi\downarrow \end{cases}. \quad (4.12)$$

де δ_{th}^{suit} - об'єктивно прийнятний рівень порогу.

Отже, у будь-якому випадку точність обробки за умов використання деякого фіксованого $\overline{\delta_{th}}$ порогу зменшуватиметься.

У свою чергу, підхід, у рамках якого використовується адаптивний поріг δ_{th}, величину δ_{th} встановлює відносно максимального рівня $\delta_{th}^{(max)}$. Для цього попередньо виконується оцінка величин δ_1 та δ_2 фрагментів $f_{x,y}^{(F)}$ на рівні кадру, тайлу або слайсу. Далі величина порогу δ_{th} нормується відповідно до виразу:

$$\delta_{th} := \delta_{th}^{(max)} \mid \delta_{th} \geq \sigma\delta_{th}^{(max)}, \quad (4.13)$$

де σ - множник, величина якого може знаходитися у діапазону [0,5;1).

Зрозуміло, що шляхом налаштування відповідного рівня σ можливо збільшувати показник точності Ф обробки аналогічно виразу (4.12), а саме:

$$\begin{cases} \sigma\downarrow \rightarrow D\uparrow, R\downarrow \rightarrow \Phi\downarrow \\ \sigma\uparrow \rightarrow D\downarrow, R\uparrow \rightarrow \Phi\downarrow \end{cases}. \quad (4.14)$$

Таким чином, досліджуваний метод може забезпечити можливість налаштування рівня достовірності.

У свою чергу, методи виявлення контурів, що базуються на просторовій фільтрації, також у сутності є пороговими. При цьому, порогові значення може бути встановлено як для модулю $|\nabla\delta|$ вектора градієнта, так і для кута χ.

Незалежно від того, який саме операнд може бути використано для обробки, існує можливість налаштування величин порогу для зазначених величин.

Таким чином, для даної групи методів також може забезпечуватися можливість налаштування величини достовірності обробки.

4.1.3 Порівняння підходів до виявлення контурів за функціональними відмінностями

Методи просторової фільтрації, спрямовані на виявлення контурів, початково орієнтовані на обробку зображень, як цілісних об’єктів.

У той же час, порогові методи виявлення факту присутності контурів демонструють ефективність у випадку поділу зображень на локальні ділянки однакового розміру. Такі методи, відповідно, є з самого початку адаптованими для обробки фрагментів $f_{x,y}^{(F)}$ кадру з використанням базису MPEG.

З іншого боку, методи просторової фільтрації можуть здійснювати обробку зображень довільного розміру $k \times \ell$.

При цьому також важливим є те, що методи просторової фільтрації можуть виконувати обхід зображення у ході виявлення контурів за такими алгоритмами, як [19, 21]:

- алгоритм повного порядкового сканування;
- алгоритм контурного обходу (алгоритм «жук») (рис.4.2).

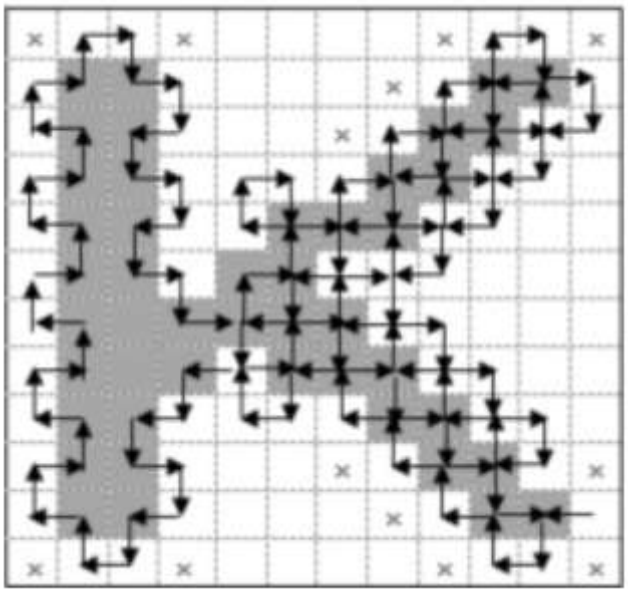


Рисунок 4.2 – Принцип обходу зображення у ході виявлення контуру за алгоритмом «жук»

Для випадку першого алгоритму обробці підлягають усі $k \times \ell$ пікселів довільного зображення, або $n \times n$ - для обробки фрагментів $f_{x,y}^{(F)}$.

У той же час, залежно від особливостей контуру та наявності примикаючих сторонніх контурів фактична кількість пікселів μ, які необхідно буде обробити для його виявлення та локалізації, може бути суттєво нижчою ніж $k \times \ell$, тобто, $\mu < k \times \ell$.

Проте, розбиття цілісного зображення $k \times \ell$ на фрагменти $n \times n$ ускладнює процес обходу контурів, оскільки єдиний контур, який розміщується у межах деякої кількості фрагментів $f_{x,y}^{(F)}$, виявляється при цьому як сукупність незалежних контурів. За рахунок цього складаються умови для суттєвого збільшення кількості E обчислювальних операцій, що еквівалентно виразу:

$$E \to E_{max}, \tag{4.15}$$

де $E_{max} \le 900 + (m+4) \times 64$, як попередньо було розраховано.

Підсумкові результати порівнянь методів виявлення контурів за критеріями швидкодії, функціональних відмінностей та достовірності демонструє табл.4.1

Таблиця 4.1 - Порівняння методів виявлення контурів за ключовими показниками

Метод / Показник	Порогові методи виявлення факту присутності контуру	Методи просторової фільтрації (на базі операнду Собеля)
Швидкодія за найгірших умов	558 операцій на фрагмент	$900 + (m+4) \times 64$ операцій на фрагмент
Достовірність	Рівень корегується	Рівень корегується
Функціональні особливості	Адаптовані для застосування на рівні фрагментів кадру	Зростає кількість обчислювальних операцій зі збільшенням рівня докомпозиції кадру, на якому виконується обробка

Таким чином, з аналізу таблиці 4.1 бачимо, що попри можливість корегування рівню достовірності, методи просторової фільтрації характеризуються суттєво нижчою швидкодією порівняно з методами виявлення факту присутності контурів на деяких локальних ділянках.

При цьому, вищі результати за показником швидкодії для методів просторової фільтрації забезпечуються у ході обробки зображень, як єдиної структури.

Отже, для використання у якості інструменту пошуку семантично-значимих ділянок відеокадрів більш прийнятною є група порогових методів виявлення факту наявності контурів.

Означена група методів характеризується:

- вищою швидкодією порівняно з методами просторової фільтрації;
- фактично незалежністю швидкодії від особливості змісту оброблюваних фрагментів відеокадрів;
- адаптованістю до схеми MPEG.

ВИСНОВКИ

У процесі виконання кваліфікаційної роботи, відповідно то умов технічного завдання, було виконано:

1. Розглянуто роль відеоінформації у розвитку сучасного соціуму
2. Досліджено загальні принципи функціонування сучасних технологій відеокодування та можливості, що надаються ними. У результаті цього виявлено, що:

- принцип функціонування будь-яких технологій кодування базується на усуненні однієї а частіше – кількох типів надмірностей;
- переважний відсоток відеоінформації у світі кодується з використанням технологій сімейства MPEG, з яких найбільш розповсюдженими зараз є група технологій H.26*;
- застосування технологій сімейства MPEG забезпечує стабільно високий коефіцієнт стиснення та прийнятний рівень якості;
- особливості змісту відеокадрів а також специфіка схеми MPEG є причиною пульсуючого характеру відеотрафіку;
- застосування режиму VBR для відеоінформації у загальному випадку є невиправданим, оскільки не ураховує природу відеотрафіку.

3. Виявлено, що у складі сучасних технологій кодування відео на рівні стандарту не затверджено інструментів, які спрямовані на урахування особливостей змісту кодованих фрагментів та встановлення для них відповідної величини коефіцієнта стиснення. Це, у свою чергу, веде до нераціонального розподілу інформаційної інтенсивності між структурними одиницями відеопотоку. Як наслідок – ряд структурних одиниць відеоряду отримає необґрунтовано перебільшену кількість біт для їх опису, що йде у розріз з їхньою фактичної семантичною значимістю. Водночас – для ряду інших структурних одиниць буде виділено обсяг біт, що для збереження їх семантики може бути недостатньо.
4. Обгрунтовано необхідність збільшення точності обробки відео, що створює умови для раціонального розподілу інформаційної інтенсивності між структурними одиницями відео потоку.

5. Обгрунтовано вибір фрагменту 8х8, як ключової структурної одиниці, та субфрагменту 4х4 – як додаткової, на рівні якої виконується кодування відео.

6. Показано, що найбільш семантично значимими ділянками відео потоку на рівні кадру є контури. Відповідно, фрагменти відеокадру, що містять контурні дані, мають кодуватися окремо від інших фрагментів.

7. Для виокремлення контурних фрагментів від фрагментів, що не містять котурів, можуть використовуватися:

- прості порогові методи, що виявляють сам факт присутності контурної інформації у межах фрагменту;

- методи на базі ковзаючи масок, що виконують виявлення та локалізацію контурів у межах зображень довільного типу.

8. Виявлені контурні фрагменти кодуються зі зменшеним коефіцієнтом стиснення, тоді як фрагменти, що контурів не містять – зі збільшеним. Це створює умови для раціонального розподілу інтенсивності усередині кодованих кадрів.

9. Доведено, що використання простих порогових методів, що орієнтовані на виявлення самого факту присутності контурної інформації, є ефективнішим методів на базі ковзаючи масок за такими показниками, як:

- швидкодія алгоритму. Кількість операцій, необхідних для виявлення присутності контуру у фрагменті кадру для простиг порогових методів за найгірших умов (двопрохідна обробка) є вдвічі меншою, ніж для методів на базі ковзаючих масок (за наближеного розрахунку арктангенсу та наближеного обчислення цілочисельного квадратного кореня), а також є незмінною, що дозволяє прогнозувати обчислювальне навантаження. При цьому, методи даної групи є потенційно продуктивнішими, так як у ході обчислень використовують лише операції додавання, віднімання та порівняння;

- функціональні особливості. Прості порогові методи є початково адаптованими до обробки фрагментів 8х8 або 4х4 відеокадру, та за будь-яких умов їхня продуктивність є сталою. У свою чергу, методи на базі ковзаючих масок є залежними від особливостей змісту аналізованих фрагментів. При цьому, зі зменшенням рівня декомпозиції кадру їхня продуктивність суттєво зменшується.

Водночас, обидві групи методів забезпечують можливість налаштування точності виявлення контурів, керуючись при цьому:

- величиною фіксованого або адаптивного порогу яскравості (для простих порогових методів) ;

- величиною модуля вектору градієнту та кута градієнту (для методів на базі ковзаючих масок).

Отже, усі завдання виконано у повному об’ємі.

.

.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Digital in 2020: World’s internet users pass the 4 billion mark [Електронний ресурс] – Режим доступу: https://wearesocial.com/blog/2020/01/global-digital-report-2020
2. VNI Forecast Highlights Tool [Електронний ресурс] – Режим доступу: https://www.cisco.com/c/m/en_us/solutions/service-provider/vni-forecast-highlights.html .
3. Requirements and future of smart cities – Express Computer [Електронний ресурс] – Режим доступу: https://www.expresscomputer.in/columns/icts-and-smart-cities/2434
4. Video quality testing in mobile networks | Rohde &Schwarz [Електронний ресурс] – Режим доступу: https://www.rohde-schwarz.com/solutions/test-and-measurement/mobile-network-testing/expertise/video-quality-testing/video-quality-testing_232054.html?change_c=true
5. MPEG-4 AVC / H.264 area [Електронний ресурс] – Режим доступу: http://compression-links.info/MPEG-4_AVC_H264
6. MPEG Home Page [Електронний ресурс] – Режим доступу: https://web.archive.org/web/20060422160837/http://www.chiariglione.org/mpeg/
7. Research Overview >> Electronics Research Group [Електронний ресурс] – Режим доступу: https://erg.abdn.ac.uk/research/future-net/digital-video/mpeg2.html
8. Ватолин Д. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин. – М. : ДИАЛОГ – МИФИ, 2003. – 384 с.
9. Best Codecs and Video Processing Software, Promo Codecs and Deals [Електронний ресурс] – Режим доступу: http://neuron2.net/library/mpeg2/iso13818-1.pdf
10. Ричардсон Ян. H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia / Ян Ричардсон. – Город. : Издательство, 2005. – 368 с.
11. Гонсалес Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1073 с.

12. RTP Payload Format for JPEG-compressed Video [Електронний ресурс] – Режим доступу: https://datatracker.ietf.org/doc/html/rfc2435
13. Чемпен Н., Чемпен Д. Цифровые технологии мультимедиа. – М.: Вильямс, 2006. – 624 с.
14. Сэломон Д. Сжатие данных, изображений и звука / Д. Сэломон. – М.: Техносфера, 2004. – 368 с.
15. Айфичер Эммануил С. Цифровая обработка сигналов: практический поход / Эммануил С. Айфичер, Барри У. Джервис. – 2-е изд. – М. : Вильямс, 2008. – 992 с.
16. The SSIM Index for Image Quality Assessment [Електронний ресурс] – Режим доступу: https://ece.uwaterloo.ca/~z70wang/research/ssim/
17. DiffIMG Download | SourceForge.net [Електронний ресурс] – Режим доступу: https://sourceforge.net/projects/diffimg/
18. Дубовик В.П., Юрик І.І. Вища математика: Навчальний посібник. – К., 2005. – 648 с.
19. Анисимов Б.В. Распознавание и цифровая обработка изображений – М.: Высш. школа, 1983. – 295с.
20. Tverdokhleb, V., Kulitsa, O.,Okladnoy, D., Hahanova, A. The development method for evaluating the saturation of video frame blocks to reduce the processing time of the video stream. IEEE East-West Design and Test Symposium (EWDTS): proceedings of Intern. Symp., (Yerevan, Armenia, October 14-17, 2016). Yerevan, 2016. P. 1-3
21. Sharpening and Edge Detection [Електронний ресурс] – Режим доступу: http://fourier.eng.hmc.edu/e161/lectures/gradient/gradient.html
22. Upbyte. Калькуляторы. Выделение границы онлайн [Електронний ресурс] – Режим доступу: http://upbyte.net/news/vydelenie_granicy_onlajn/2018-11-03-1041
23. Shi, Yun Q. Image and video compression for multimedia engineering: fundamentals, algorithms, and standards / Yun Q Shi, Huifang Sun.