

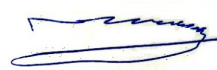
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Факультет електроніки
(повна назва інституту/факультету)

Кафедра акустичних та мультимедійних електронних систем
(повна назва кафедри)

«На правах рукопису»
УДК 621.397:004.627

«До захисту допущено»

Завідувач кафедри



Сергій НАЙДА
(ніціали, прізвище)

“12” січня 2024 р.

Магістерська дисертація

спеціальність 171 «Електроніка»
(код і назва)

на тему: «Дослідження алгоритмів стиснення даних для забезпечення ефективності передавання мультимедійної інформації»

Виконав: студент II курсу, групи ДВ-21мп
(шифр групи)

Сіренко Олексій Володимирович
(прізвище, ім'я, по батькові)


(підпис)


Керівник доцент, к.т.н., доцент Лазебний В. С.
(посада, науковий ступінь, вчене звання, прізвище та ініціали)


(підпис)

Консультант _____
(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

Рецензент доц. кафедри ЕПС доц., к.т.н. Цибульський Л.Ю
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище, ініціали)


(підпис)

Засвідчую, що у цьому дипломному проекті немає запозичень з праць інших авторів без відповідних посилань.

Студент


(підпис)

Київ – 2024 року

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

Факультет _____ електроніки _____

Кафедра _____ акустичних та мультимедійних електронних систем _____

Рівень вищої освіти – другий (магістерський) за освітньо-професійною програмою _____

Спеціальність _____
(освітня програма) _____ 171 «Електроніка» («Електронні системи мультимедіа та засоби Інтернету речей»)

ЗАТВЕРДЖУЮ
Завідувач кафедри



Сергій.НАЙДА
(ініціали, прізвище)

« 14 » грудня 2022 р.

**ЗАВДАННЯ
на магістерську дисертацію студенту**

Сіренку Олексію Володимировичу
(прізвище, ім'я, по батькові)

- 1 Тема роботи _____ «Дослідження алгоритмів стиснення даних для забезпечення ефективності передавання мультимедійної інформації» _____
- керівник роботи _____ Лазебний Володимир Семенович, к.т.н., доцент. _____
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «09» листопада 2023 р. № 5218–с

- 2 Строк подання студентом дисертації _____ 10 січня 2024 р. _____
3. Об'єкт дослідження Алгоритми стиснення даних.
4. Вхідні дані для магістерської дисертації за освітньо-професійною програмою: стандарт MPEG–2 , стандарт MPEG–4 алгоритми функціонування кодексів стиснення відеозображень
5. Перелік завдань, які потрібно розробити: проаналізувати сучасні принципи стиснення відеоданих; дослідити алгоритми стиснення зображень в системах та MPEG–2 і MPEG–4; проаналізувати принципи об'єктивного оцінювання відеозображень; зробити натурний експеримент з оцінювання якості відтвореного зображення, стиснутого із застосуванням кодеків MPEG–2 і MPEG–4.

6. Перелік графічного (ілюстративного) матеріалу: комп'ютерна презентація результатів досліджень_.

7. Консультанти розділів дисертації


Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

8. Дата видачі завдання 10. 09. 2022 р .

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Аналіз основ стиснення даних	11.09.2023 – 30.09.2023	Виконано
2	Розгляд цифрового стиснення та потоків MPEG	1.10.2023 – 15.10.2023	Виконано
3	Аналіз процесів стиснення MPEG–2	16.10.2023 – 21.10.2023	Виконано
4	Аналіз процесів стиснення H.264	21.10.2023 – 31.10.2023	Виконано
5	Об'єктивна оцінка якості сприйняття відеоматеріалу	02.11.2023 – 15.11.2023	Виконано
6	Класифікація та критерії оцінки об'єктивних методів вимірювання якості відеозображень	16.11.2023 – 30.11.2023	Виконано
	Опрацювання матеріалів до першого. Другого та третього розділів дисертації та написання висновків до них	1.12.2023 – 19.12.2023	Виконано
	Дослідження кодеків за допомогою оцінки якості зображення об'єктивним методом	1.01.2024 – 8.01.2024	Виконано
	Оформлення дисертації підготовки презентації до захисту11	8.01.2024 – 12.01.2024	Виконано

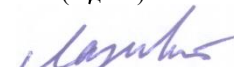
Студент


(підпис)

Олексій СІРЕНКО

(ініціали, прізвище)

Керівник роботи


(підпис)

Володимир ЛАЗЕБНИЙ

(ініціали, прізвище)

РЕФЕРАТ

Сіренко О.В. Дослідження алгоритмів стиснення даних для забезпечення ефективності передавання мультимедійної інформації: магістерська дисертація за освітньо-професійною програмою : 171 Електроніка. – Київ, 2024. – 101 с.

АРТЕФАКТИ, ЕФЕКТИВНІСТЬ, КОДЕКИ, СТИСНЕННЯ ВІДЕО, H.264, MPEG-2

Актуальність дослідження. MPEG-2 та MPEG-4 є ключовими у сфері стиснення відеоданих. MPEG-2 був розроблений для передавання зображень з високою якістю в системах цифрового телебачення та цифрового кінематографу, MPEG-4 був розроблений для передавання мультимедійної інформації в інформаційних мережах з різною пропускнуою здатністю. Широке застосування мультимедійної інформації в електронних засобах комунікацій обумовлює необхідність правильного вибору системи стиснення відео для економного використання пропускнуої здатності телекомунікаційних каналів і забезпечення необхідної якості відтвореного зображення.

Мета дослідження. Дослідження алгоритмів стиснення даних для забезпечення ефективності передавання мультимедійної інформації.

Завдання дисертаційної роботи: проаналізувати сучасні принципи стиснення відеоданих; дослідити алгоритми стиснення зображень в системах та MPEG-2 і MPEG-4; проаналізувати принципи об'єктивного оцінювання відеозображень; зробити натурний експеримент з оцінювання якості відтвореного зображення, стиснутого із застосуванням кодеків MPEG-2 і MPEG-4.

Об'єкт дослідження: Системи стиснення зображень.

Предмет дослідження: Ефективність стиснення відео та характеристики кодеків H.264 і MPEG-2.

Методи дослідження: Критичний аналіз системи MPEG-2 та системи MPEG-4, порівняльний аналіз алгоритмів стиснення відео реалізованих в MPEG-2 та MPEG-4, натурний експеримент для оцінювання об'єктивних показників якості стиснених зображень.

Новизна отриманих результатів: Отримано числові значення об'єктивних показників якості стиснених зображень для оцінювання ефективності функціонування кодеків MPEG-2 та MPEG-4.

Практичне значення результатів: Результати можуть бути використані при виборі системи стиснення для каналів з різною пропускнуою здатністю, а також для відеосистем зі спеціальними вимогами до якості відтвореного зображення.

ABSTRACT

Sirenko O.V. Research of data compression algorithms to ensure the efficiency of multimedia information transmission: master's thesis under the educational and professional program: 171 Electronics. – Kyiv, 2024. – 101 p.

ARTIFACTS, PERFORMANCE, CODECS, VIDEO COMPRESSION, H.264, MPEG-2

Relevance of research. MPEG-2 and MPEG-4 are key in video compression. MPEG-2 was developed for transmission of images with high quality in digital television and digital cinema systems, MPEG-4 was developed for transmission of multimedia information in information networks with different bandwidth. The widespread use of multimedia information in electronic means of communication determines the need for the correct selection of a video compression system for economical use of the bandwidth of telecommunication channels and ensuring the necessary quality of the reproduced image.

The aim of the study. Research of data compression algorithms to ensure the efficiency of multimedia information transmission.

The task of the dissertation: to analyze the modern principles of video data compression; to investigate image compression algorithms in MPEG-2 and MPEG-4 systems; analyze the principles of objective evaluation of video images; conduct a real-life experiment to evaluate the quality of the reproduced image compressed using MPEG-2 and MPEG-4 codecs.

Research object: Image compression systems.

Research subject: Video compression efficiency and characteristics of H.264 and MPEG-2 codecs.

Research methods: Critical analysis of MPEG-2 system and MPEG-4 system, comparative analysis of video compression algorithms implemented in MPEG-2 and MPEG-4, field experiment for evaluation of objective quality indicators of compressed images.

Novelty of the obtained results: Numerical values of objective indicators of the quality of compressed images were obtained for evaluating the effectiveness of MPEG-2 and MPEG-4 codecs.

Practical significance of the results: The results can be used when choosing a compression system for channels with different bandwidths, as well as for video systems with special requirements for the quality of the reproduced image.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	9
ВСТУП.....	10
1 СУЧАСНІ ПРИНЦИПИ СТИСНЕННЯ ВІДЕОЗОБРАЖЕНЬ.....	12
1.1 Значення стиснення.....	12
1.2 Принципи стиснення.....	12
1.3 Стиснення зображень в телебаченні.....	13
1.4 Вступ до цифрового стиснення відео.....	15
1.5 Потоки MPEG.....	19
1.6 Потреба в моніторингу та аналізі.....	21
1.7 Особливості стиснення з втратами.....	22
Висновки до розділу.....	24
2 ДОСЛІДЖЕННЯ АЛГОРИТМІВ СТИСНЕННЯ ЗОБРАЖЕНЬ В СИСТЕМАХ ТА MPEG–2 І MPEG–4.....	25
2.1 Еволюція стандартів	25
2.2 Продуктивність відео	26
2.3. Формування та передавання цифрових потоків за стандартом MPEG–2.....	27
2.4 Стандарт H.264.....	51
Висновки до розділу.....	66
3 ПРИНЦИПИ ОБ’ЄКТИВНОЇ ОЦІНКИ ОТРИМАНОГО ВІДЕОМАТЕРІАЛУ.....	67
3.1 Об’єктивна оцінка якості сприйняття відеоматеріалу.....	67
3.2 Пікове відношення сигнал/шум.....	67
3.3 Середнє пікове співвідношення сигнал/шум.....	68
3.4 Середньоквадратична помилка.....	69
3.5 Середньоквадратична абсолютна різниця.....	70
3.6 Середньоквадратична різниця.....	71
3.7 Індекс структурної схожості зображення.....	72

3.8 Міра якості відео на основі дискретного косинусного перетворення	76
3.9 Алгоритм визначення якості зображення на основі методу ледь помітних відмінностей Сарнова.....	78
3.10 Інші методи об'єктивного вимірювання якості зображень.....	79
3.11 Класифікація та критерії оцінки об'єктивних методів вимірювання якості відеозображень.....	83
Висновки до розділу	85
4 ДОСЛІДЖЕННЯ КОДЕКІВ ЗА ДОПОМОГОЮ ОБ'ЄКТИВНИХ МЕТОДІВ ВИМІРЮВАННЯ ЯКОСТІ ВІДЕОЗОБРАЖЕНЬ.....	86
4.1 Опис програмного забезпечення для вимірювання якості	87
4.2 Постановка експерименту дослідження кодеків за допомогою оцінки якості зображення об'єктивним методом.....	88
Висновки до розділу.....	93
ВИСНОВКИ	96
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	97
ДОДАТОК.....	99

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

- ПЕП – Packetized Elementary Stream (пакетизований елементарний потік);
- PAL – Phase–Alternating Line, європейська система аналогового кольорового телебачення;
- MPEG – Motion Picture Experts Group, експертна група з рухомих зображень;
- ISO – International Organization for Standardization, Міжнародна організація стандартизації;
- ВП – відеопослідовність;
- ДКП – дискретне косинусне перетворення;
- AVC – Advanced Video Coding, покращене кодування відеозображення;
- MSAD – Mean Square Absolute Difference, середньоквадратична абсолютна різниця;
- MSSSIM – Multi–Scale Structure Similarity Index Measure, багатомасштабний індекс структурної схожості зображення;
- NE – Noise Estimation Metric, метрика оцінки рівня шуму;
- PQR – Picture Quality Rating, оцінка якості зображення;
- PSNR – Peak Signal to Noise Ratio, пікове відношення сигнал/шум;
- SCD – Scene Change Detector, метрика моментів зміни сцени;
- SSIM – Structure Similarity Index Measure, індекс структурної схожості зображення;
- VQM – Video Quality Measurement, вимірювання якості відео;
- SPTS – Single Program Transport Stream, програмний транспортний потік;
- PAT – Program Association Table, таблиця асоціацій програм;
- PMT – Program Map Table, таблиць мапи програм;
- MSE – Mean Squared Error, середньоквадратична помилка;
- APSNR – Average Peak Signal–to–Noise Ratio, середнє пікове співвідношення сигнал/шум;
- NQI – Non–reference Quality Index, індекс якості без посилань

ВСТУП

В епоху високошвидкісного Інтернету та широкого використання цифрових технологій, передача мультимедійної інформації стає ключовою складовою сучасного способу життя. З кожним днем зростає популярність онлайн–стрімінгу, відеоконференцій, і інших мультимедійних платформ, що ставить перед вченими і інженерами завдання забезпечити не тільки стабільність передачі даних, але й високу якість відтворення відео.

У цьому контексті актуальність дослідження алгоритмів стиснення даних набуває особливого значення. Застосування ефективних методів стиснення, зокрема стосовно відеоданих, стає необхідністю для забезпечення оптимального використання мережевих ресурсів і задоволення зростаючих вимог до якості відеоконтенту.

Актуальність дослідження. MPEG–2 та MPEG–4 є ключовими у сфері стиснення відеоданих. MPEG–2 був розроблений для передавання зображень з високою якістю в системах цифрового телебачення та цифрового кінематографу, MPEG–4 був розроблений для передавання мультимедійної інформації в інформаційних мережах з різною пропускнуою здатністю. Широке застосування мультимедійної інформації в електронних засобах комунікацій обумовлює необхідність правильного вибору системи стиснення відео для економного використання пропускнуої здатності телекомунікаційних каналів і забезпечення необхідної якості відтвореного зображення.

Мета дослідження. Дослідження алгоритмів стиснення даних для забезпечення ефективності передавання мультимедійної інформації.

Завдання дисертаційної роботи:

- проаналізувати сучасні принципи стиснення відеоданих;
- дослідити алгоритми стиснення зображень в системах та MPEG–2 і MPEG–4;
- поаналізувати принципи об’єктивного оцінювання відеозображень;
- зробити натурний експеримент з оцінювання якості відтвореного зображення, стиснутого із застосуванням кодеків MPEG–2 і MPEG–4.

Об'єкт дослідження: Системи стиснення зображень.

Предмет дослідження: Ефективність стиснення відео та характеристики кодеків MPEG-4 і MPEG-2.

Методи дослідження: Критичний аналіз системи MPEG-2 та системи MPEG-4, порівняльний аналіз алгоритмів стиснення відео реалізованих в MPEG-2 та MPEG-4, натурний експеримент для оцінювання об'єктивних показників якості стиснених зображень.

Новизна отриманих результатів: Отримано числові значення об'єктивних показників якості стиснених зображень для оцінювання ефективності функціонування кодеків MPEG-2 та MPEG-4.

Практичне значення результатів: Результати можуть бути використані при виборі системи стиснення для каналів з різною пропускнуою здатністю, а також для відеосистем зі спеціальними вимогами до якості відтвореного зображення.

1 СУЧАСНІ ПРИНЦИПИ СТИСНЕННЯ ВІДЕОЗОБРАЖЕНЬ

1.1 Значення стиснення

Спочатку успіх цифрового відео пов'язаний із застосуванням на етапі постпродакшну, де висока вартість цифрового відео компенсується його безмежними можливостями для накладання шарів та ефектів. Однак цифрове відео стандарту виробництва генерує понад 200 мегабіт на секунду, а така швидкість передачі даних вимагає великої ємності для зберігання і широкої смуги пропускання для передачі. Цифрове відео могло б знаходити застосування в більш широких областях, якби вимоги до зберігання та пропускну здатності були зменшені, і ось для цього існує стиснення.

Стиснення – це метод вираження цифрового аудіо та відео за допомогою меншого обсягу даних. Стиснення має наступні переваги:

- Для заданої кількості вихідного матеріалу потрібен менший обсяг сховища.
- При роботі в режимі реального часу стиснення зменшує необхідну пропускну здатність. Крім того, стиснення дозволяє передавати дані між носіями швидше, ніж у реальному часі, наприклад, між стрічкою та диском. Для заданої кількості вихідного матеріалу потрібен менший обсяг сховища.
- Стислий формат запису може використовувати меншу щільність запису, зроблюючи диктофон менш чутливим до факторів навколишнього середовища та технічного обслуговування.

1.2 Принципи стиснення

Існують два принципово різні методи, які можна використовувати для зменшення кількості даних, що передають інформаційний контент. У практичних системах стиснення, вони зазвичай поєднуються, часто дуже складним чином.

Перший спосіб – це підвищення ефективності кодування. Існує багато способів кодування будь-якої інформації, і більшість простих представлень даних

відео та аудіо містять значну кількість надлишковості. Поняття ентропії розглядається нижче. Для зменшення або усунення надмірності можна використовувати багато прийомів кодування; приклади включають кодування довжини виконання та системи кодування змінної довжини, такі як коди Хаффмана. При правильному використанні ці методи є повністю оборотними, так що після декомпресії дані будуть ідентичні тим, що були на вході системи. Цей тип стиснення відомий як стиснення без втрат. Такі комп'ютерні програми для архівації, як PKZip, використовують стиснення без втрат.

Очевидно, що стиснення без втрат є ідеальним, але, на жаль, воно зазвичай не забезпечує необхідного ступеня зменшення даних для відео– та аудіододатків. Однак, оскільки стиснення відбувається без втрат, воно може застосовуватися в будь–якій точці системи і часто використовується на виході даних компресорів з втратами.

Якщо усунення надмірності не зменшує дані настільки, наскільки це необхідно, деяку інформацію доведеться відкинути. Системи стиснення з втратами досягають зменшення обсягу даних, видаляючи інформацію, яка не є релевантною або менш важливою. Це не загальні методи, які можна застосувати до будь–якого потоку даних; оцінку релевантності можна зробити лише в контексті програми, розуміючи, що представляють дані і як вони будуть використовуватися. У випадку телебачення, застосуванням є презентація зображень і звуку для зорових і слухових систем людини, і для розробки ефективної системи стиснення необхідно добре розуміти людський фактор. Деяка інформація у відеосигналах не може бути сприйнята зоровою системою людини і, отже, є дійсно несуттєвою в даному контексті. Система стиснення, яка відкидає лише несуттєву інформацію про зображення, називається візуально без втрат.

1.3 Стиснення зображень в телебаченні

Телевізійні сигнали, будь то аналогові чи цифрові, завжди представляли значну кількість інформації, і методи зменшення пропускної здатності

використовувалися вже на ранніх етапах. Ймовірно, одним з перших прикладів є використання черезрядкової розгортки. За фіксованої кількості рядків і частоти оновлення зображення, черезрядкова розгортка дозволяє зниження пропускну здатності в 2:1. Цей процес пов'язаний із втратами, адже черезрядкова розгортка викликає артефакти, спричинені інтерференцією між вертикальною та часовою інформацією, і зменшує корисну вертикальну роздільну здатність зображення. Тим не менше, більшість відмовлившихся аспектів були неважливими, і черезрядкова розгортка стала простим і цінним компромісом. На жаль, цей метод та артефакти, що виникають, стають складними для обробки у більш сучасних схемах цифрового стиснення. Значна складність MPEG-2 пов'язана з обробкою черезрядкових сигналів, і ефективність кодування значно зменшується порівняно з прогресивними сигналами.

Подальший розвиток відбувся з введенням кольору. Кольорові камери генерують GBR-сигнали, які теоретично містять втричі більше інформації, ніж монохромний сигнал. Однак потрібно було передавати кольорові сигнали тими ж каналами, що і монохромні.

Рішення частково полягало в перетворенні GBR-сигналів на сигнал яскравості (означається Y), а також двох сигналів різниці кольорів, U і V (або I і Q). Створення сигналу яскравості дозволило вирішити проблему сумісності з монохромними приймачами. Важливим кроком для зменшення пропускну здатності було фільтрування сигналів різниці кольорів, оскільки зорова система людини майже не реагує на високочастотну інформацію в сигналах різниці кольорів. Останнім етапом у процесі зменшення пропускну здатності стало "приховування" сигналів різниці кольорів у невикористаних частинах спектра монохромного сигналу.

Деякі цифрові методи можуть використовувати схожі принципи, такі як MPEG, яка перетворює сигнали для виділення несуттєвої інформації. Фільтрація сигналів різниці кольорів також залишається важливою, але цифрові технології дозволяють здійснювати більш ефективно фільтрування для економії пропускну здатності.

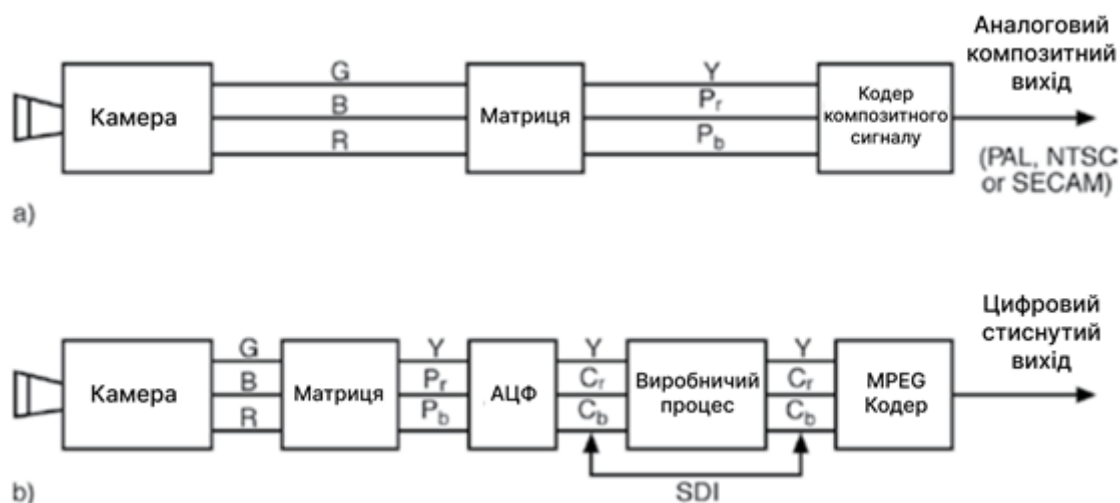


Рисунок 1.1 – Традиційні та сучасні телевізійні системи

На рис.1.1 а) показано, що в традиційних телевізійних системах сигнал камери GBR перетворюється на компоненти Y , P_b , P_r для виробництва і кодується в аналоговий композитний сигнал для передачі.

На рис. 1.1 б) показано сучасний еквівалент. Сигнали Y , P_b , P_r оцифровуються і передаються як сигнали Y , C_b , C_r у формі SDI (Цифровий інтерфейс, який може використовуватися для передачі цифрового відеосигналу) через виробничий процес перед кодуванням за допомогою MPEG для подальшої передачі. Очевидно, що MPEG може розглядатися як більш ефективна заміна композитного відео. Крім того, MPEG має більшу гнучкість, оскільки необхідний бітрейт можна регулювати відповідно до застосування. З меншими бітрейтами і роздільною здатністю, MPEG можна використовувати для відеоконференцій і відеотелефонів.

1.4 Вступ до цифрового стиснення відео

У будь-якому реальному програмному матеріалі можна виділити два типи компонентів сигналу: ті, які є новими і непередбачуваними, і ті, які можна передбачити. Новий компонент називається ентропією і є справжньою інформацією в сигналі. Решта називається надмірністю, оскільки вона не має великого значення. Надмірність може бути просторовою, наприклад, на великих рівнинних ділянках зображення, де сусідні пікселі мають майже однакові

значення. Також надмірність може бути часовою, коли існує схожість між послідовними зображеннями. Усі системи стиснення працюють, відокремлюючи ентропію від надмірності в кодері. Тільки ентропія записується або передається, а декодер обчислює надмірність з переданого сигналу. Рис. 1.2 а) демонструє цю концепцію. Ідеальний кодер виділяє всю ентропію, і лише вона передається декодеру. Проте на практиці досягнення цього ідеалу неможливо.

Ідеальний кодер виявився б складним і спричинив би значну затримку в обробці часової надмірності. У деяких додатках, таких як запис або мовлення, певна затримка є прийнятною, але в відеоконференціях вона є непринятною. В деяких випадках дуже складний кодер також може бути занадто витратним. Таким чином, не існує однієї ідеальної системи стиснення.

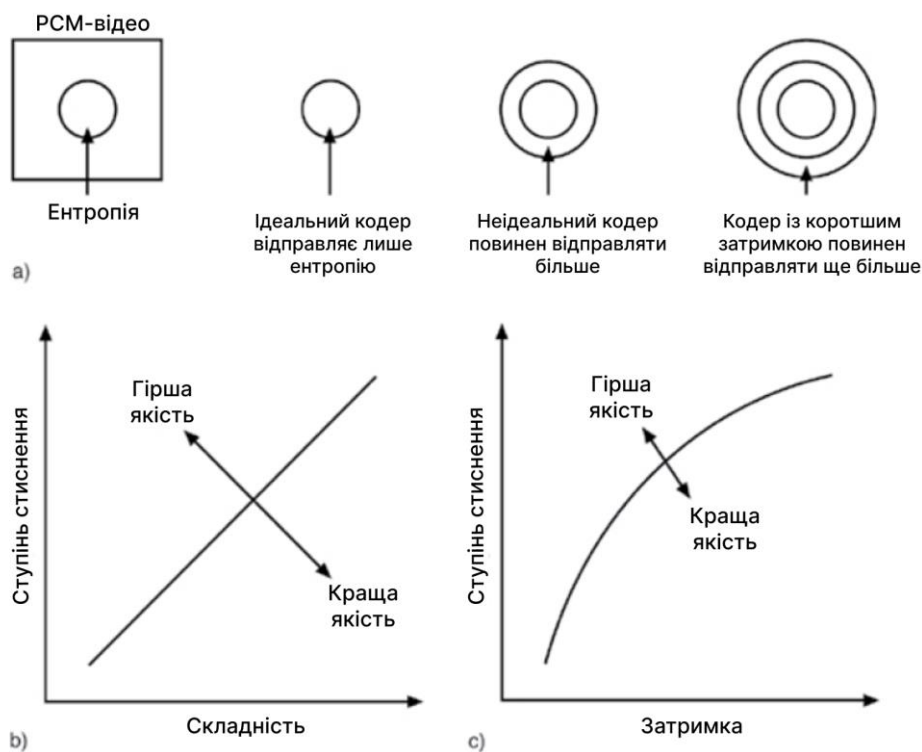


Рисунок 1.2 – Компроміс між якістю, складністю та затримкою

На практиці необхідно використовувати цілий ряд кодерів з різними затримками та складністю обробки. Однією з переваг MPEG є те, що це не єдиний формат стиснення, а складається з цілої низки стандартизованих інструментів кодування, які можна комбінувати для різних застосувань. Спосіб кодування міститься у стиснутих даних, щоб декодер міг автоматично впоратися з тим, що

вирішив реалізувати кодер.

У MPEG-2 і MPEG-4 кодування поділено на кілька профілів з різною складністю, і кожен профіль може бути використаний на різних рівнях, залежно від роздільної здатності вхідного зображення. Розділ 4 надає докладну інформацію щодо профілів та рівнів.

Існує безліч різних цифрових відеоформатів, кожен з яких має свою власну швидкість передачі даних. Наприклад, система високої чіткості може мати швидкість передачі даних у шість разів більшу, ніж система стандартної чіткості.

Отже, просто знати бітрейт, який надає кодер, не завжди є корисним. Важливим є коефіцієнт стиснення, тобто відношення вхідного бітрейту до стисненого, наприклад, 2:1, 5:1 і так далі. Однак кількість задіяних змінних робить дуже складним визначення відповідного коефіцієнта стиснення.

На рис. 1.2 а) показано, що для ідеального кодера, якщо вся ентропія передається, якість є високою. Однак зі збільшенням коефіцієнта стиснення для зниження бітрейту не вся ентропія передається, і якість падає. У стисненій системі, коли відбувається втрата якості, ця втрата є значною (рис. 1.2, б). Якщо доступна швидкість передачі даних недостатня, краще уникнути цього етапу, зменшивши ентропію вхідного зображення. Це можна зробити за допомогою фільтрації. Втрата роздільної здатності, спричинена фільтрацією, суб'єктивно є більш прийнятною, ніж артефакти стиснення. Щоб ідеально визначити ентропію, ідеальний компресор мав би бути надзвичайно складним. Практичний компресор може бути менш складним з економічних міркувань і повинен надсилати більше даних, щоб бути впевненим у тому, що він передає всю ентропію. На рис. 1.2 б) показано зв'язок між складністю кодера та його продуктивністю. Чим вищий необхідний коефіцієнт стиснення, тим складнішим повинен бути кодер.

Ентропія відеосигналів може змінюватися. Наприклад, запис диктора, який представляє новини, має велику кількість надмірної інформації і легко стискається. З іншого боку, важко стиснути запис із колючого листя, що коливається на вітрі, або запис футбольного натопу, який постійно рухається і, отже, має менше надмірності (більше інформації або ентропії). У будь-якому

випадку, якщо не передати всю ентропію, це призведе до втрати якості. Таким чином, можна вибирати між каналом із постійною швидкістю передачі даних із змінною якістю або каналом із постійною якістю із змінною швидкістю передачі даних. Оператори телекомунікаційних мереж, як правило, віддають перевагу постійній швидкості передачі даних з практичних міркувань, але буферна пам'ять може бути використана для усереднення варіацій ентропії, якщо збільшення затримки, що виникає в результаті, є прийнятним. Під час запису із змінною швидкістю передачі даних може бути легше працювати, і DVD використовує змінну швидкість передачі даних, використовуючи буферизацію, щоб середня швидкість передачі даних залишалася в межах можливостей дискової системи.

Інтракодування (intra = всередині) – це техніка, яка використовує просторову надмірність або надмірність всередині зображення; інтеркодування (inter = між) – це техніка, яка використовує часову надмірність. Інтракодування може використовуватися окремо, як у стандарті JPEG для нерухомих зображень, або в поєднанні з інтеркодуванням, як у MPEG.

Інтракодування спирається на дві характеристики типових зображень. По-перше, не всі просторові частоти присутні одночасно, а по-друге, чим вища просторова частота, тим менша амплітуда. Інтракодування вимагає аналізу просторових частот у зображенні. Цей аналіз є метою таких перетворень, як вейвлети та ДКП (дискретне косинусне перетворення). Перетворення створюють коефіцієнти, які описують величину кожної просторової частоти. Зазвичай, багато з цих коефіцієнтів будуть нульовими або майже нульовими, і їх можна опустити, що призведе до зменшення швидкості передачі даних.

Інтракодування ґрунтується на пошуку подібностей між послідовними зображеннями. Коли декодер має доступ до конкретного зображення, наступне може бути сформоване шляхом передачі лише відмінностей між зображеннями. Різниця між зображеннями збільшується, коли об'єкти рухаються, але це збільшення можна компенсувати за допомогою компенсації руху, оскільки рухливий об'єкт, як правило, слабо змінює свій вигляд від одного зображення до іншого. Якщо рух може бути вимірний, більш точне наближення до поточного

зображення можна створити, переміщуючи частину попереднього зображення на нове місце. Процес переміщення контролюється парою значень горизонтального та вертикального зміщення (разом відомих як вектор руху), які передаються на декодер. Передача вектора руху вимагає менше даних, ніж передача інформації про різницю між зображеннями.

MPEG може обробляти як черезрядкові, так і не черезрядкові зображення. Зображення в певний момент часу називається "кадром", незалежно від того, чи це поле чи кадр. Черезрядкова розгортка не є ідеальним джерелом для цифрового стиснення, оскільки це, в суті, техніка стиснення. Часове кодування ускладнюється тим, що пікселі в одному полі розташовані в іншому полі інакше, ніж пікселі в іншому.

Компенсація руху мінімізує, але не усуває відмінності між послідовними зображеннями. Сама різниця між зображеннями є просторовим зображенням і може бути стиснута за допомогою внутрішнього кодування на основі перетворень, як описано раніше. Компенсація руху просто зменшує кількість даних у різницевому зображенні.

Ефективність часового кодера зростає зі збільшенням проміжку часу, протягом якого він може діяти. Як показано на рис. 1.2 с), якщо необхідний високий коефіцієнт стиснення, слід враховувати довший часовий проміжок на вході і, відповідно, більшу затримку кодування. Очевидно, що часові кодовані сигнали важко редагувати, оскільки вміст вихідного зображення може базуватися на даних зображення, переданого деякий час тому. Виробничі системи повинні обмежувати ступінь часового кодування, щоб забезпечити можливість редагування, а це обмеження, в свою чергу, обмежить доступний коефіцієнт стиснення.

1.5 Потіки MPEG

Вихідний сигнал одного аудіо або відеокодера MPEG відомий як елементарний потік. Елементарний потік – це нескінченний сигнал, що дуже

близький до реального часу. Для зручності елементарний потік можна розділити на блоки даних з контрольованим розміром, створюючи пакетований елементарний потік (ПЕП). Ці блоки даних вимагають заголовкової інформації для ідентифікації початку пакетів і повинні містити мітки часу, оскільки пакування порушує часовий характер.

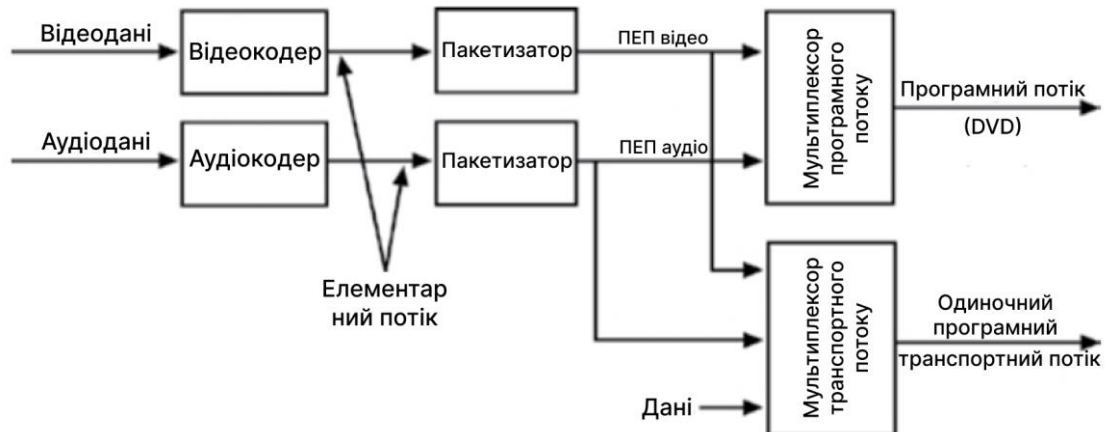


Рисунок 1.3 – Етапи кодування та інкапсуляції при створенні програмних та транспортних потоків

На рис. 1.3 показано, що один відео ПЕП та кілька аудіо ПЕП можуть бути об'єднані в програмний потік, за умови, що всі кодеки працюють за спільним годинником. Мітки часу в кожному ПЕП можна використовувати для забезпечення синхронізації між відео та аудіо. Потоки програм мають пакети змінної довжини зі заголовками. Вони використовуються для передачі даних на оптичні та жорсткі диски, де очікується низький рівень помилок, і де файли можуть мати різний розмір. DVD використовує програмні потоки.

Для передачі та цифрового мовлення кілька програм і пов'язаних з ними ПЕП можуть бути мультиплексовані в один транспортний потік.

Транспортний потік відрізняється від програмного потоку тим, що пакети ПЕП поділяються на короткі пакети фіксованого розміру і що можна передавати кілька програм, закодованих з різними тактовими частотами. Це можливо завдяки тому, що транспортний потік має механізм еталонного програмного тактового генератора (PCR), який дозволяє передавати декілька тактових частот, одну з яких

вибирається і регенерується на декодері. Також можливий єдиний програмний транспортний потік (SPTS), який розташовується між кодером і мультиплексором. Оскільки транспортний потік може синхронізувати тактові сигнали декодера з тактовими сигналами кодера, SPTS є більш поширеним, ніж програмний потік.

Транспортний потік – це більше, ніж просто мультиплекс аудіо та відео ПЕП. Окрім стиснених аудіо-, відео- та даних, транспортний потік включає метадані, що описують бітовий потік. Сюди входить таблиця асоціацій програм (PAT), в якій перераховані всі програми в транспортному потоці. Кожен запис у PAT вказує на таблицю мапи програм (PMT), яка перелічує елементарні потоки, з яких складається кожна програма. Деякі програми будуть відкритими, але до деяких може бути застосовано умовний доступ (шифрування), і ця інформація також міститься у метаданих. Транспортний потік складається з пакетів даних фіксованого розміру, кожен з яких містить 188 байт. Кожен пакет містить код ідентифікатора програми (PID). Пакети в одному елементарному потоці мають однаковий PID, так що декодер (або демультіплексор) може вибрати потрібний елементарний потік і відкинути решту. Підрахунок безперервності пакетів гарантує, що кожен пакет, необхідний для декодування потоку, буде отримано. Ефективна система синхронізації необхідна для того, щоб декодер міг правильно ідентифікувати початок кожного пакета і десеріалізувати бітовий потік у слова.

1.6 Потреба в моніторингу та дослідженні

Транспортний потік MPEG є надзвичайно складною структурою, що використовує взаємопов'язані таблиці та кодовані ідентифікатори для розділення програм та елементарних потоків всередині програм. Усередині кожного елементарного потоку існує складна структура, що дозволяє декодеру розрізняти, наприклад, вектори, коефіцієнти та таблиці квантування.

Збої можна розділити на дві великі категорії. У першій категорії транспортна система правильно передає інформацію від кодера/мультиплексора до декодера без бітових помилок або додаткового джиттера, але

кодер/мультиплексор або декодер має несправність. У другій категорії кодер/мультиплексор і декодер справні, але несправна передача даних від одного до іншого. Дуже важливо знати, де саме виникла несправність – в кодері/мультиплексорі, транспортній системі чи декодері, щоб знайти швидке рішення.

Проблеми синхронізації, такі як втрата або пошкодження шаблонів синхронізації, можуть перешкоджати отриманню всього транспортного потоку. Дефекти протоколу транспортного потоку можуть завадити декодеру знайти всі дані для програми, можливо, передаючи зображення, але не звук. Коректна доставка даних, але з надмірним джиттером може спричинити проблеми з синхронізацією декодера.

Якщо система, що використовує транспортний потік MPEG, виходить з ладу, несправність може бути в кодері, мультиплексорі або декодері. Як можна ізолювати цю несправність? По–перше, перевірте, чи відповідає транспортний потік стандартам кодування MPEG. Якщо потік не відповідає стандартам, навряд чи можна звинувачувати декодер у виникненні труднощів. Якщо ж потік відповідає стандартам, можливо, декодер потребує уваги.

Традиційні інструменти для тестування відео – генератор сигналів, монітор форми сигналу і векторний мікроскоп – не підходять для аналізу систем MPEG, за винятком випадків, коли потрібно переконатися, що відеосигнали, які входять і виходять з системи MPEG, мають відповідну якість. Натомість для тестування приймального обладнання та декодерів необхідне надійне джерело достовірних тестових сигналів MPEG. За допомогою відповідного аналізатора можна з високим ступенем достовірності оцінити продуктивність кодерів, систем передачі, мультиплексорів і демультіплексорів.

1.7 Особливості стиснення з втратами

Стиснення MPEG веде до втрат, оскільки декодований файл не є ідентичним оригіналу. Ентропія джерела змінюється, і коли ентропія висока, система стиснення може залишити видимі артефакти при декодуванні. При

часовому стисненні передбачається надлишковість між послідовними зображеннями. Якщо цього немає, система може дати збій. Прикладом може бути відео з прес-конференції, де стріляють зі спалахів. Окремі зображення, що містять спалах, повністю відрізняються від своїх сусідів, і артефакти кодування можуть стати очевидними.

Нерівномірний рух або кілька об'єктів, що рухаються незалежно один від одного на екрані, вимагають великої пропускної здатності вектора, і цю вимогу можна задовольнити, лише зменшивши пропускну здатність, доступну для даних зображення. Знову ж таки, можуть виникати видимі артефакти, рівень яких змінюється і залежить від руху. Ця проблема часто виникає у відео зі спортивних репортажів.

Грубе квантування призводить до контурування яскравості та плакатного кольору. Це можна побачити як плямисті тіні та блокування на великих ділянках чистого кольору. Суб'єктивно, артефакти стиснення є більш дратівливими, ніж відносно постійні погіршення, спричинені аналоговими телевізійними системами передачі.

Єдиним рішенням цих проблем є зменшення коефіцієнта стиснення. Отже, користувачеві стиснення доводиться робити ціннісне судження між економічністю високого коефіцієнта стиснення та рівнем артефактів.

Окрім збільшення затримки кодування і декодування, тимчасове кодування також спричиняє труднощі при редагуванні. Насправді, потік бітів MPEG не може бути довільно відредагований. Це обмеження виникає тому, що при часовому кодуванні для декодування одного зображення може знадобитися вміст попереднього зображення, який може бути недоступним після редагування. Той факт, що зображення можуть надсилатися поза послідовністю, також ускладнює редагування.

Якщо було використано відповідне кодування, редагування може відбуватися, але тільки в точках з'єднання, які відносно широко рознесені. Якщо потрібне довільне редагування, потік MPEG повинен пройти процес декодування-модифікації-рекодування, що призведе до втрати генерації.

Висновки до розділу

Аналізуючи технічні аспекти стиснення, можна виділити кілька ключових проблем, які виникають під час декодування стиснених файлів.

Однією з проблем є втрати, пов'язані зі стисненням, які призводять до різних артефактів та втрат якості відновленого відео. Особливо важливою є проблема артефактів, пов'язаних із часовим стисненням, де надлишковість між послідовними кадрами може впливати на якість відтворення, особливо в ситуаціях зі стрілянням або інтенсивним рухом на екрані.

Також важливою проблемою є нерівномірний рух та обробка кількох незалежних об'єктів на екрані, що може призводити до артефактів, залежних від руху. Це особливо актуально для відео зі спортивних подій, де об'єкти можуть рухатися не прогнозовано та швидко.

Крім того, грубе квантування призводить до контурування яскравості та плакатного кольору, що стає помітним у вигляді плямистих тіней та блокування на великих ділянках чистого кольору. Ці артефакти є суттєвими з точки зору сприйняття якості відтвореного відео.

Для досягнення мети дисертаційної роботи необхідно виконати такі дослідження:

- дослідити алгоритми стиснення зображень в системах та MPEG-2 і MPEG-4;
- проаналізувати принципи об'єктивного оцінювання відеозображень;
- зробити натурний експеримент з оцінювання якості відтвореного зображення, стиснутого із застосуванням кодеків MPEG-2 і MPEG-4.

2 ДОСЛІДЖЕННЯ АЛГОРИТМІВ СТИСНЕННЯ ЗОБРАЖЕНЬ В СИСТЕМАХ ТА MPEG-2 І MPEG-4

2.1 Еволюція стандартів

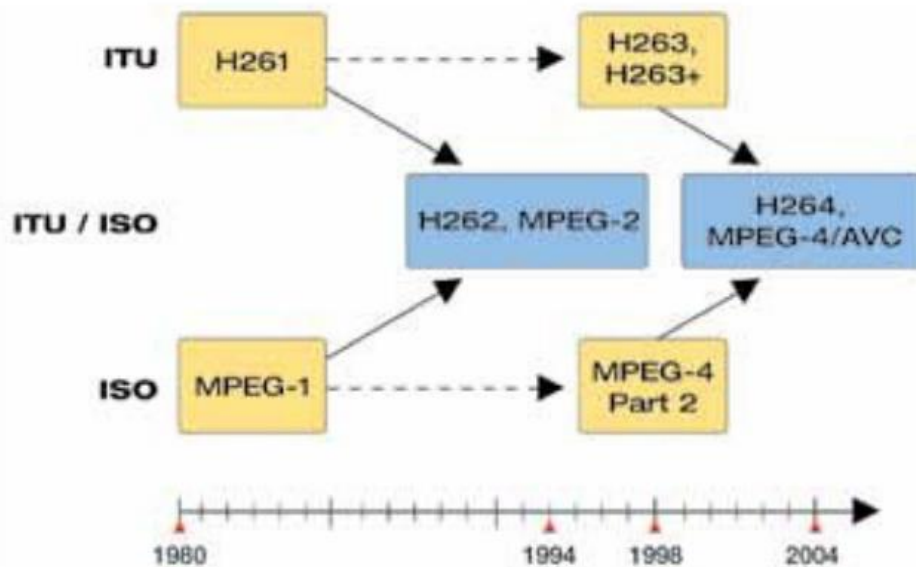


Рисунок 2.1 – Еволюція MPEG

На рис. 2.1 зображено еволюцію відеостандартів. Слід відзначити, що відеостандарт MPEG-4, частина 2, представляє собою обширний стандарт, який і досі має свої проблеми для комерційної реалізації. Він виявився занадто складним та містить численні профілі, більшість з яких не знаходять застосування в комерційних програмах. Синтаксис не завжди є настільки структурованим та чітким, що ускладнює впровадження та взаємодію. Деякі його елементи залишаються відкритими для інтерпретації, а існуючі помилки, такі як розходження між стандартом, нормативним програмним забезпеченням і бітовими потоками, часом призводять до суперечок. За великий період часу, необхідний для розробки та стандартизації MPEG-4, технологія визріла і перейшла до H.264/AVC, який безперечно надає краще стиснення.

H.264 – це назва, пов'язана з лінійкою відеостандартів H.26x MCE-T. AVC вказує на сторону ISO/IEC MPEG партнерського проекту, яка завершила стандартизацію після попереднього розроблення в рамках проекту H.26L в MCE-T. Його також можна називати H.264/ MPEG-4 AVC або MPEG-4 Part 10 для

підкреслення спільної спадщини. Іноді його називають "кодек JVT CODEC" за посиланням на організацію JVT, яка його розробила.

H.264/AVC фокусується виключно на відео, тому його стандартизація була легшою та швидшею, ніж частина MPEG-4. Крім того, він не потребує взаємодії з численними іншими частинами того ж стандарту. Розробники H.264/AVC також врахували інтеграцію його елементів у інші системи. Зокрема, з невеликими модифікаціями транспортного потоку MPEG-2 він може передавати відео H.264/AVC так само, як і відео MPEG-2. Це робить H.264/AVC дуже привабливим для галузі телерадіомовлення, що використовує транспортний потік MPEG-2 ('TS'). Щоб оновити до покращеної якості відео та HD, достатньо оновити лише відеоеlementи на кінцевих точках джерела та призначення.

Нещодавно компанія Microsoft внесла альтернативний відеокодек для AVC/H.264 на ринок мовлення. Спочатку Microsoft розробила формат Windows Media Video (WMV) як власний кодек для потокового мовлення в Інтернеті. У версії 9 WMV було доповнено новим розширеним профілем для програм мовлення. У 2003 році WMV-9 було представлено для стандартизації в SMPTE. Стандарт отримав офіційне затвердження в березні 2006 року як SMPTE 421M, відомий також як VC-1.

VC-1 надає продуктивність, подібну до H.264/AVC, і можливо вимагає менше обробки та пам'яті. Його стандартизація в SMPTE робить його потенційно широко застосовувемим. Важливо відзначити, що конкуруючі формати наступного покоління для стандарту DVD, HD DVD і Blu-ray, вимагають підтримки AVC/H.264 і VC-1.

2.2 Продуктивність відео

Продуктивність кодека – це складний показник, що визначає, скільки біт (рис. 2.2) потрібно для досягнення певної візуальної якості відеопослідовності. Проте, виміряти його набагато складніше, оскільки визначення "візуальної якості" є суб'єктивним результатом багатьох факторів, таких як:

- Різкість відео.
- Виявлення візуальних артефактів, таких як "блочність" або "дзвінкість".
- Точність передачі кольору.
- Синхронізація відео/аудіо.
- Плавність руху.

Ці візуальні особливості залежать від різних типів кодування в послідовності, різних типів кадрів та розподілу використання бітів протягом послідовності. Хоча суб'єктивні елементи важко вимірювати, об'єктивні аспекти, такі як кодування та використання бітів, можна кількісно оцінити. Існує певна кореляція між візуальною якістю та кількісними параметрами, такими як вимірювання PSNR (пікове співвідношення сигнал/шум).

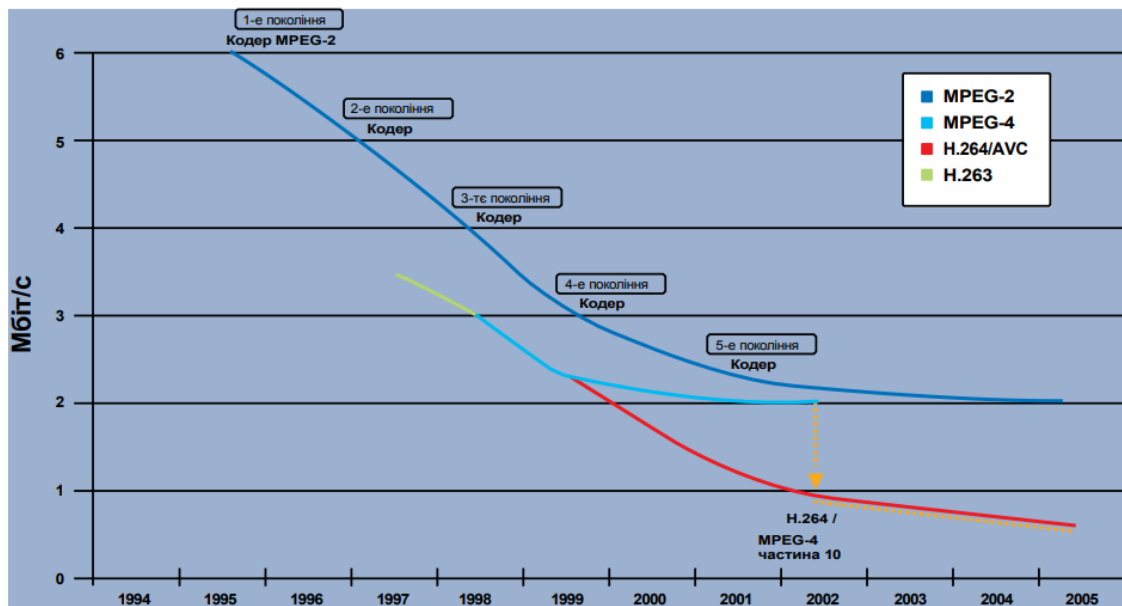


Рисунок 2.2 – Вимоги до пропускної здатності знижуються в міру вдосконалення кодеків з часом

2.3 Формування та передавання цифрових потоків за стандартом MPEG-2

2.3.1 Стиснення відеоданих

Насамперед, слід зазначити, що в офіційних документах, в яких описано процес стиснення, переважає підхід фахівців з інформатики та систем зв'язку, а

застосовувану термінологію також запозичено, головним чином, із цих сфер техніки.

Ключовим словосполученням є «синтаксис потоку», тобто суто формальний опис алгоритму та його параметрів. Такий «дедуктивний» підхід, що полягає в описі процесу від найвищого (системного) рівня, багато в чому протилежний традиційному для відео– та звуко– інженерів «індуктивного» підходу, за якого послідовно описують всі операції оброблення сигналу – «від входу до виходу». Надалі ми будемо дотримуватись легшого для сприйняття та традиційного для відеотехніки підходу.

До власне функцій стиснення не належать такі функції кодера MPEG як кодовахист, пакетування даних, об'єднання (мультиплексування) отриманих потоків даних з іншими потоками такої самої, або іншої природи, обмеження доступу до певних типів даних тощо.

Ці функції є характерними для передавання конкретних видів даних, тому має сенс розглядати їх окремо.

2.3.2 Основні складові процесу стиснення відеоданих за системою MPEG 2

Процес стиснення відеоданих можна умовно розділити на чотири етапи:

1. Колориметричне перетворення сигналів R, G, B в Y, Cb, Cr (тільки в разі використання вхідних даних у форматі R, G, B) та повторна дискретизація колірних різницевих сигналів Cb, Cr.

Для перетворення сигналів RGB у сигнали YCbCr застосовують стандартні колориметричні коефіцієнти та рівні квантування за Рекомендацією 601, хоча стандартом передбачено й інші варіанти. Роздільну здатність сигналів Cb, Cr на етапі цьому залишають незмінною або зменшують: удвічі чи в чотири рази лише по горизонталі, або по вертикалі і горизонталі одночасно.

2. Прогнозування кадру на підґрунті попередніх та/або наступних кадрів з локальною оцінкою руху окремих блоків пікселів та компенсацією руху. Цей етап

зазвичай називають зменшенням чи усуненням надлишковості в часі (запобігають передаванню інформації про вже відомі пікселі в наступних кадрах).

3. Двовимірне Дискретне Косинусне Перетворення та адаптивне квантування (більш грубе) отриманих цифрових значень коефіцієнтів ДКП. Саме на цьому етапі виникають найістотніші спотворення зображення (але ці спотворення практично не помітні глядачу).

4. Статистичне (ентропійне) кодування призначених значень кодом Хафмана.

У результаті такого кодування жодних спотворень зображення не виникає, а обсяг даних суттєво зменшується. Сукупно третій та четвертий етапи реалізують усунення просторової надлишковості.

Керування швидкістю потоку стислих відеоданих на виході кодера (наприклад, для підтримки постійної швидкості цього потоку) здійснюють трьома способами:

- шляхом зміни шкал квантування, тобто, зрештою, шляхом зниження або підвищення якості декодованого зображення;
- шляхом зміни порядку сканування коефіцієнтів ДКП для регулювання ефективності ентропійного кодування;
- шляхом введення «порожніх» даних додаткового потоку (stuffing), який ніяк не впливає на якість декодованого зображення, але дозволяє скоригувати надто малу швидкість потоку.

Узагальнені схеми процесів кодування та декодування за системою МПЕГ 2 наведено на рис.2.3 та рис. 2.4.

Застосуванням вищезазначених способів, розробники кодерів стиснення прагнуть забезпечити як якість зображення окремих кадрів чи фрагментів відеопрограми, так і однорідність (consistency) цієї якості для всіх кадрів відеоряду. Завдання є надзвичайно не простим. Справа в тому, що глядач досить спокійно ставиться до загальної втрати якості, наприклад, до втрати чіткості діагональних рухомих структур, але вкрай неприємними для сприйняття є часті зміни якості, наприклад, різкі зміни контрасту діагональних структур (мерехтіння

і тремтіння). Саме тут проходить лінія вододілу між низькоякісним та високоякісним обладнанням для стиснення.

На відміну від кодера для декодера MPEG не має бути свободи прийняття рішень. Декодер лише реалізує всі передбачені стандартом функції перетворення даних відповідно до одержуваних від кодера інструкцій, але в зворотному порядку щодо оброблення даних у кодері.

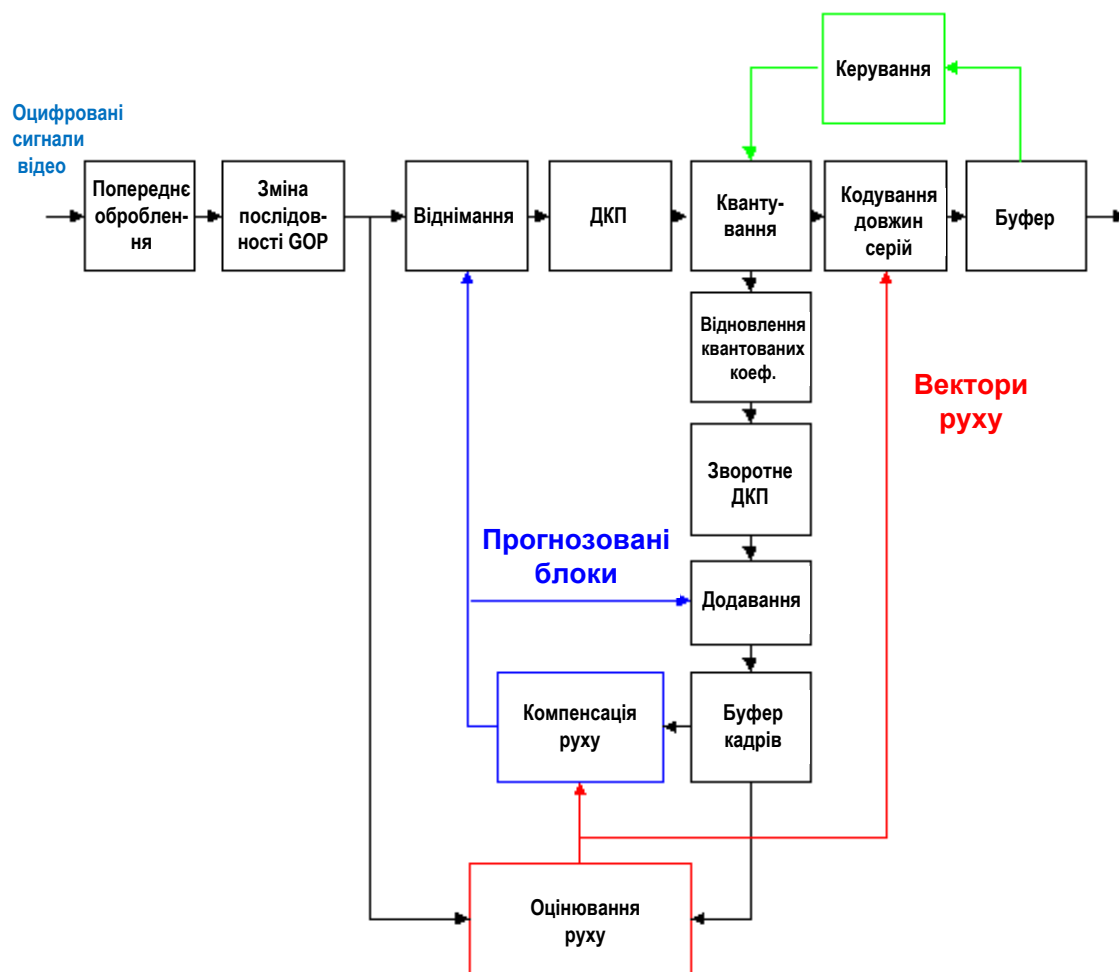


Рисунок 2.3 – Блок–схема типового кодера стиснення відеоданих

Декодер реалізує такі функції:

- декодування кодів Хафмана (відновлення кодів фіксованої довжини);
- відновлення сигналу за квантованими відліками (зворотне квантування), що відновлює округлені значення коефіцієнтів ДКП повного діапазону;
- зворотне Дискретне Косинусне Перетворення, що приблизно відновлює значення параметрів пікселів;

– реконструкція повних решіток дискретизації всіх відеокадрів із компенсацією руху та записом отриманих значень пікселів за правильними адресами до буфера кадрової пам'яті;

– зворотне колориметричне перетворення Y, C_b, C_r в R, G, B (опційне та композитне кодування)

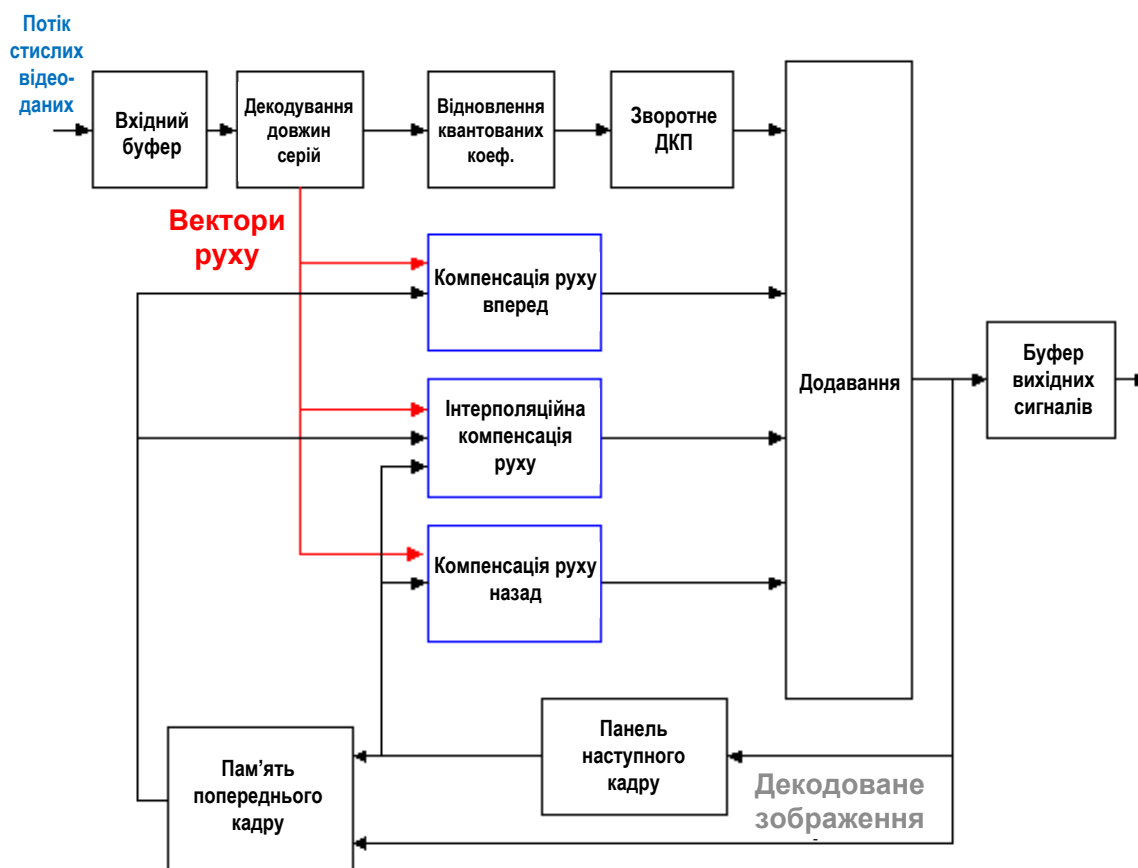


Рисунок 2.4 – Типова блок–схема декодування стислих відеоданих

2.3.3 Усунення часової надлишковості та прогнозування відеокадрів

Розглянемо просторово–часове оброблення сигналів зображення у кодері MPEG–2.

Як впливає з рис. 2.5, початковою інформацією для всього процесу стиснення є три прямокутні матриці (таблиці) відліків сигналів RGB або YC_bCr (у європейському інженерному жаргоні замість правильного позначення YC_bCr дуже часто використовують успадковане від системи PAL позначення YUV).

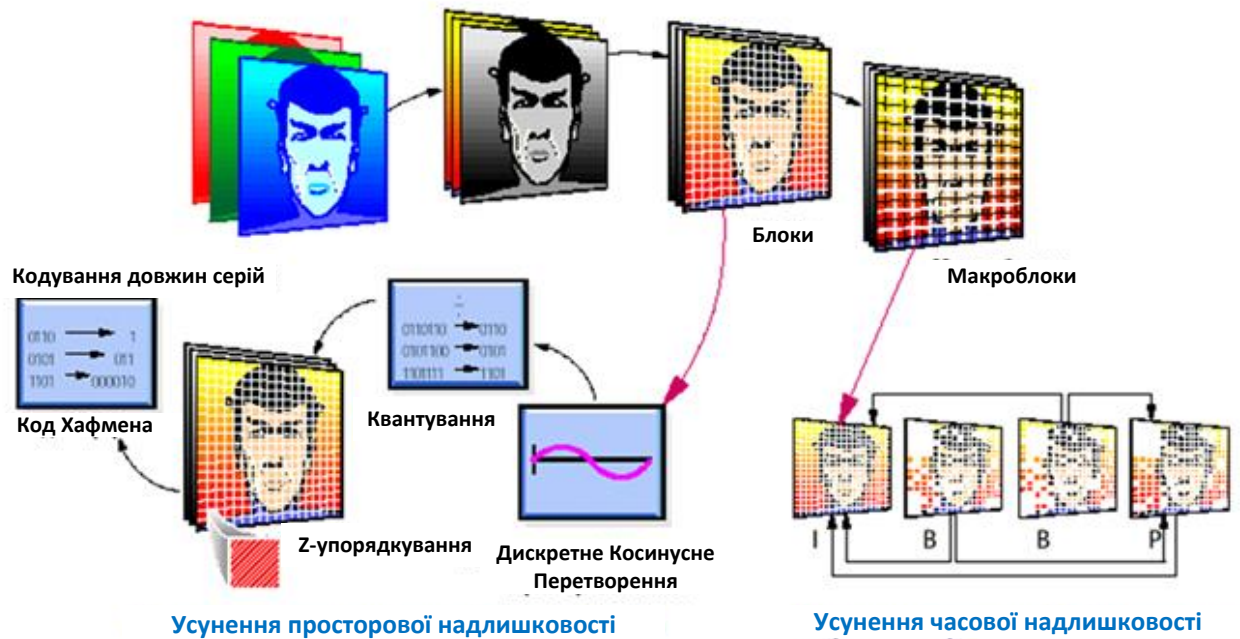


Рисунок 2.5 – Схема просторово–часового оброблення під час стиснення відеоданих

Розміри матриць Y та C_b , C_r можуть бути як однаковими, так і різними. Передбачено досить велику кількість варіантів розмірів і співвідношень розмірів цих матриць, що передбачено у стандарті. Справа в тому, що стандартом MPEG–2 передбачено не тільки передавання зображень мовленнєвого телебачення, але і багато інших застосунків. Однак, у будь–якому разі розміри матриці по горизонталі і вертикалі мають бути кратними 16. Це пов'язано з поблочним перетворенням відліків зображення методом Дискретного Косинусного Перетворення (Discrete Cosine Transform). Навіть, якщо розмір початкового зображення не є кратним 16, то для цілей подальшого оброблення його завжди доповнюють «порожніми» пікселями до кратного розміру.

Один з основних принципів стиснення – це передавання лише інформації, яка змінюється від кадру до кадру, рис.2.6. Міжкадрове прогнозування є мало ефективним для оброблення сцен, де багато рухомих об'єктів. Компенсація руху дозволяє більш точно прогнозувати ті частини зображення, де рух (переміщення відносно розташування в попередньому кадрі) можна виміряти, рис.2.7.

Якщо рух виміряно досить точно, тоді для кожного макроблоку елементів зображення (розміром 16 x 16 елементів) можна передавати лише відносно малі міжкадрові відмінності плюс двовимірний вектор руху, рис.2.8. Слід підкреслити, що в такому разі помилки вимірювання вектора руху не призводять до помітних спотворень зображення, а лише знижують ефективність стиснення, тобто. призводять до зростання переданого сигнального потоку.

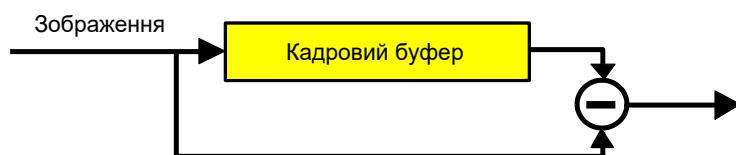


Рисунок 2.6 – Схема обчислення міжкадрових поелементних різниць для прогнозування кадру



Рисунок 2.7 – Схема процесу прогнозування кадру з компенсацією руху

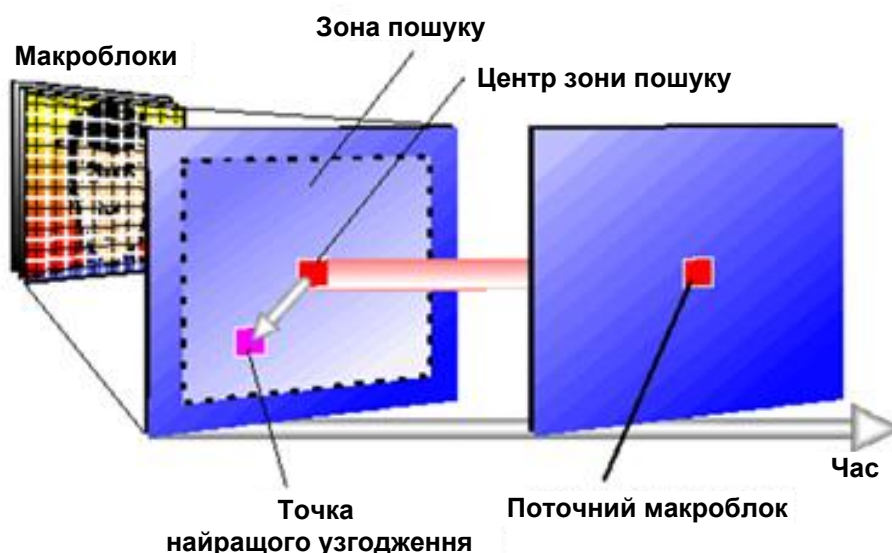


Рисунок 2.8 – Ілюстрація до процесу обчислення вектору руху

Процедури оцінювання й компенсації руху належать до найскладніших.

Стандартом передбачено чотири варіанти компенсації руху: кадрова компенсація (frame motion compensation), польова компенсація (field motion compensation), внутрішньопольова компенсація «16x8» і двоступінчаста компенсація з уточненням вектора (dual prime motion compensation). Неякісна (спрощена) оцінка, отже, і неповна компенсація руху, можуть призводити до зростання обсягу стислих даних загалом на 15–20 %.

2.3.4 I-, P- та B-кадри

Декодер MPEG не здатний відновити всі відеокадри довгої послідовності лише на підґрунті інформації про міжкадрові різниці. Час від часу, необхідно надсилати до декодера стислий (з усуненням просторової надлишковості), але не прогнозований, а повний відеокадр, або хоча б макроблок (без усунення часової надлишковості). Такий кадр або макроблок позначають латинською літерою I (Intra – внутрішній).

Деякі кадри можуть бути добре прогнозовані за попередньою інформацією. Ці кадри позначають літерою P (Predicted). Для формування (прогнозування) поточного відеокадру типу P використовують найближчий попередній кадр (типу I або P).

У межах кадру типу B передбачено можливість передавати макроблоки восьми підтипів – прогнозовані з компенсацією руху, прогнозовані без компенсації руху, взагалі прогнозовані (типу I), і навіть пропущені (skipped). Макроблоки підтипу «пропущений» має сенс передавати, наприклад, коли міжкадрові різниці не перевищують рівня шумів (на однорідних ділянках статичного зображення).

Кадри, розташовані між кадрами типу I і P, можуть бути прогнозовані як із застосуванням інформації попередніх кадрів, так і майбутніх кадрів, рис.2.9.

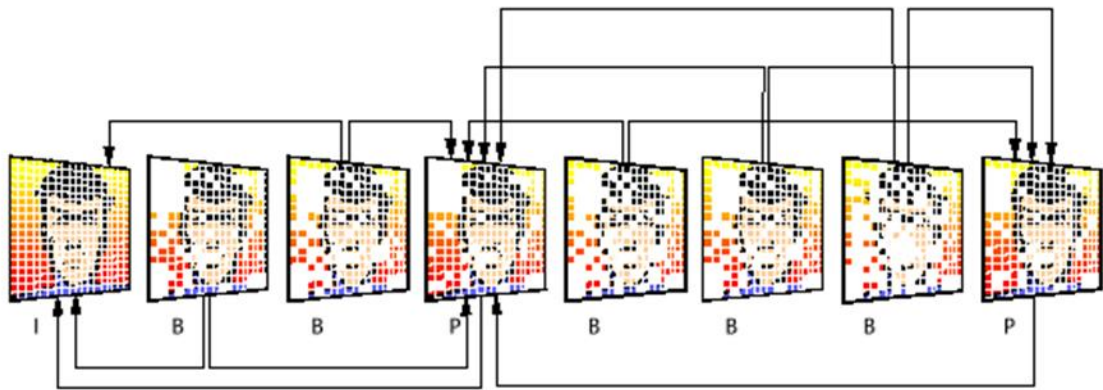


Рисунок 2.9 – Приклад структури групи кадрів

Такі кадри позначають літерою В (Bi-directionally predicted) – двобічно прогнозовані. Кадри типу В не використовують як підґрунтя для прогнозування інших кадрів. У межах кадру типу В можна передавати макроблоки дванадцяти підтипів – з «прогнозом вперед», «прогнозом назад» і з двосторонньою інтерполяцією.

Декілька відеокadrів типу I, P і В об'єднують в групу, яку називають GOP (Group Of Pictures) – група картин. Розмір та структуру групи не регламентовано стандартом, однак, на практиці, у багатьох застосунках передбачено застосовувати фіксовані структури GOP. Наприклад, для досягнення заданої суб'єктивної якості зображення із застосуванням структур I, I, I, I... треба використовувати швидкості бітового потоку в 2,5–3 рази більші, ніж у разі використання структури I, B, B, P...

Очевидно, що інтерполяція навіть порівняно точно прогнозованих, але далеко віддалених у часі кадрів навряд чи призведе до хороших результатів. Тому розробники систем прагнуть обмежити максимальну відстань часової інтерполяції для того, щоб вона не перевищувала 3–4 кадрових інтервалів.

Запропоновано позначати структуру групи кадрів із застосуванням двох чисел M і N. Число N позначає загальну кількість кадрів у групі: від першого кадру (зазвичай – кадру типу I) до останнього кадру (типу P або B), що передуює початковому кадру наступної групи. Число M означає максимальну кількість кадрів B, розташованих між прогнозованими кадрами типу P.

Для телевізійного мовлення найчастіше застосовують структуру з $M=3$ $N=12$, яку іноді позначають просто (3,12).

У процесі передавання групи відеокадрів порядок їхнього слідування не обов'язково має співпадати з початковим порядком. Навпаки, для успішної часової інтерполяції необхідно, щоб до моменту обчислення поточного кадру (типу В або Р) всі кадри, що використані для його створення, мають бути наявними в пам'яті декодера. Призначення кодера – подавати інформацію так, щоб виконати зазначену вище умову. Згідно із зазначеною тезою послідовність наведена на рис.2.9, що має структуру із восьми кадрів I1 B2 B3 P4 B5 B6 B7 P8 має надійти на вхід декодера у такому порядку: I1 P4 B2 B3 P8 B5 B6 B7. Цифра позначає порядковий номер кадру в початковій послідовності.

Для відновлення початкового порядку проходження відеокадрів декодер здійснює їх перегрупування із застосуванням відповідних додаткових сигналів – часових міток.

Рис. 2.9 – це лише приклад. Стандартом MPEG–2 передбачено можливість застосовувати практично будь–які угруповання відеокадрів. Зокрема, можлива група з одного кадру – типу I. Хоча стиснення даних у такому разі незначне, однак цей відеопотік легко монтувати. Нагадаємо, що в межах кадрів типу Р та В окремі макроблоки можуть мати інший тип, зокрема – тип I.

Розмір групи може бути змінений «на ходу». Зокрема, корисно розпочинати нову групу в момент зміни сюжету й деяких інших випадках, де компенсація руху є неефективною.

2.3.5 Усунення просторової надлишковості та Дискретне Косинусне Перетворення

Усунення просторової надлишковості, що називають також «внутрішньо–кадровим обробленням» (Intra–Frame Coding), має місце, в основному, на рівні блоку, рис.2.10.

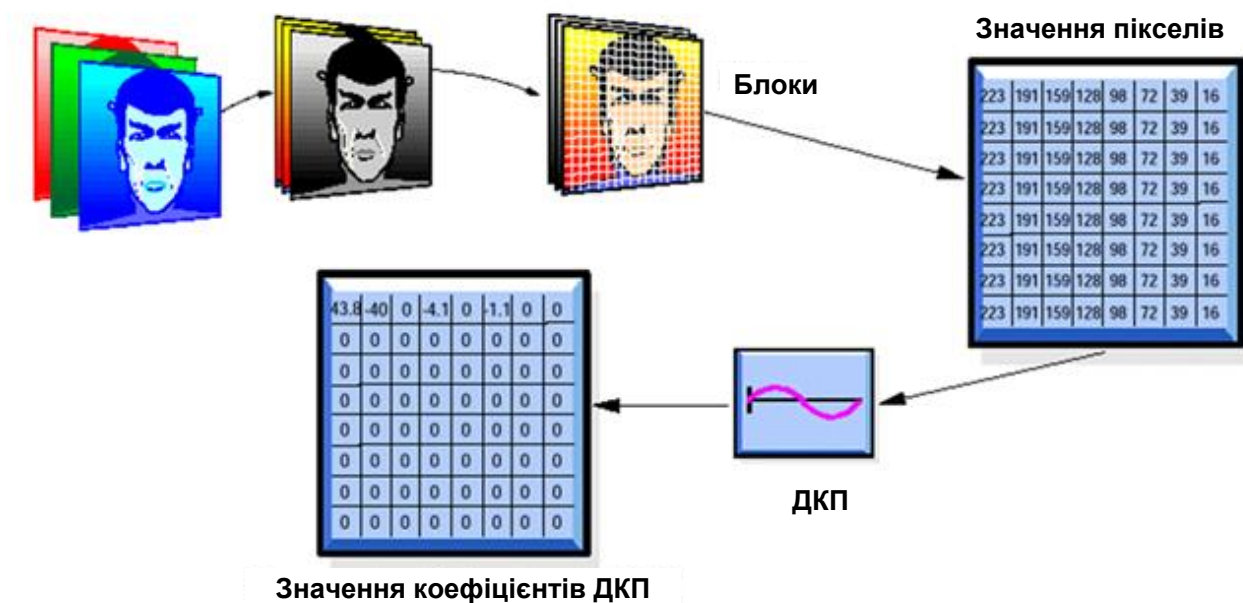


Рисунок 2.10 – Схема усунення просторової надлишковості

Термін «блок» у цьому контексті може означати як групу відліків вхідного, різницевого або декодованого зображення, так і групу відповідних коефіцієнтів ДКП. Розмір блоку яскравості фіксований: 8×8 , тобто. для кожного блоку обчислюють та передають 64 коефіцієнти двовимірного ДКП, рис.2.10.

Формально, одновимірне ДКП є результатом одностороннього Дискретного Перетворення Фур'є (ДПФ) вхідного блоку відліків, стикованого з дзеркальним відображенням цього ж блоку, рис. 2.11. Необхідної швидкості обчислень досягнуто завдяки застосуванню добре відомих алгоритмів і схемо-технічних рішень, спочатку розроблених для Швидкого Перетворення Фур'є (БПФ).

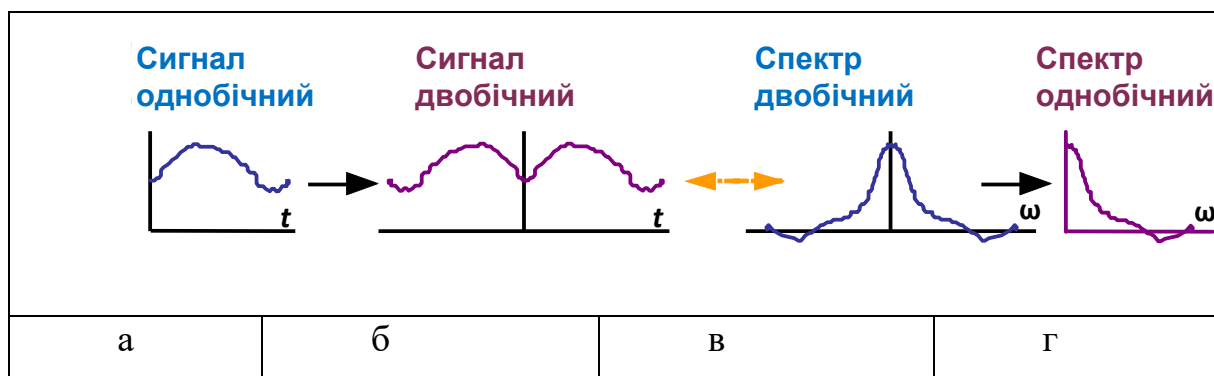


Рисунок 2.11 – ДКП є окремим різновидом ДПФ

Деталізуємо наведену вище тезу. Зауважимо, що сигнальна функція (наприклад, розподіл яскравості в певному напрямку) є випадковою залежністю розподілу яскравості пікселів від координати. Якщо застосувати дискретне перетворення Фур'є до такої функції, то отримуємо коефіцієнти для синусних і косинусних ортогональних функцій.

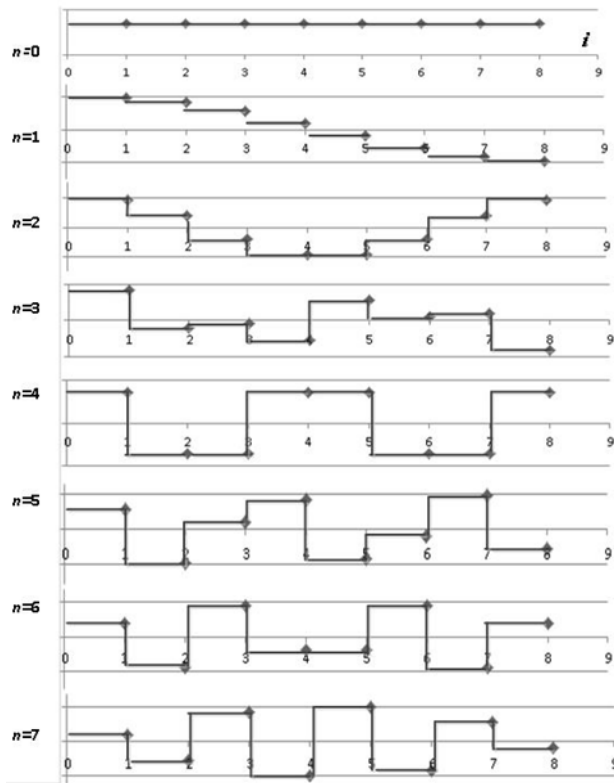


Рисунок 2.12 – Сигналограми дискретного 8-точкового косинусного перетворення Фур'є

З математики відомо, що для парних функцій (симетричних відносно осі координат, рис.2.11 б), результат перетворення містить лише значення коефіцієнтів косинусних функцій, тоді як амплітуди (коефіцієнти перетворення) всіх синусних складових взаємно компенсуються і дорівнюють нулю. Щоб можна було застосовувати ДКП для будь-якої сигнальної функції, рис.2.11 а), її спочатку доповнили дзеркальним відображенням відносно осі ординат і отримали парну функцію, рис.2.11б. Для цієї парної функції знайшли розрахункове співвідношення для визначення коефіцієнтів перетворення на подвійному інтервалі (інтервал додатних і від'ємних значень аргументів, рис.2.11 б). Знайдене співвідношення і є ДКП, в результаті якого знаходимо коефіцієнти перетворення

або іншими словами – амплітуди відповідних косинусних базисних функцій.

На рис. 2.12 наведено зображення ортогональних функцій 8–точкового дискретного Косинусного перетворення (без урахування нормувального коефіцієнта $\sqrt{(2/N)}$), n – номер базисної функції, i – номер відліку дискретизованої функції.

Точні значення відліків 8–точкового перетворення з урахуванням нормувальних коефіцієнтів наведено в табл.2.1.

Таблиця 2.1 – Значення відліків 8–точкового ДКП

	n								
	= 0	1	2	3	4	5	6	7	
i	0	0	0	0	0	0	0	0	0
=0	,35	,49	,46	,415	,355	,28	,19	,1	
1	0	0	0	–	–	–	–	–	–
1	,35	,415	,19	0,1	0,355	0,49	0,46	0,28	
2	0	0	–	–	–	0	0	0	0
2	,35	,28	0,19	0,049	0,39	,1	,46	,415	
3	0	0	–	–	0	0	–	–	–
3	,35	,1	0,46	0,28	,355	,415	0,19	0,49	
4	0	–	–	0	0	–	–	–	0
4	,35	0,1	0,46	,28	,355	0,415	0,19	,49	
5	0	–	–	0	–	–	0	–	–
5	,35	0,28	0,19	,049	0,355	0,1	,46	0,415	
6	0	–	0	0	–	0	–	–	0
6	,35	0,415	,19	,1	0,355	,49	0,46	,28	
7	0	–	0	–	0	–	0	–	–
7	,35	0,49	,46	0,415	,355	0,28	,19	0,1	

Доцільно згадати таку властивість дискретного перетворення із застосуванням ортогональних функцій:

– дискретне перетворення Фур'є функції, визначеної N рівномірно взятими

відліками (відліки від 0 до $N - 1$) є періодичним з періодом N ;

– кількість ортогональних функцій для здійснення дискретного перетворення Фур'є для сигнальної функції, що містить N відліків з рівномірним кроком дорівнює N .

Тобто, для того, щоб описати довільну дискретизовану сигнальну функцію можна застосувати значення відліків (амплітуди) в N точках на осі аргументу або N коефіцієнтів перетворення Фур'є для ортогональних функцій, визначених на тому ж відрізку, що і сигнальна функція.

2.3.6 Двовимірне Дискретне Косинусне Перетворення

Двовимірне ДКП – це комбінація двох одновимірних перетворень ДКП, здійснених в горизонтальному та вертикальному напрямках (більш узагальнено – за двома ортогональними напрямками).

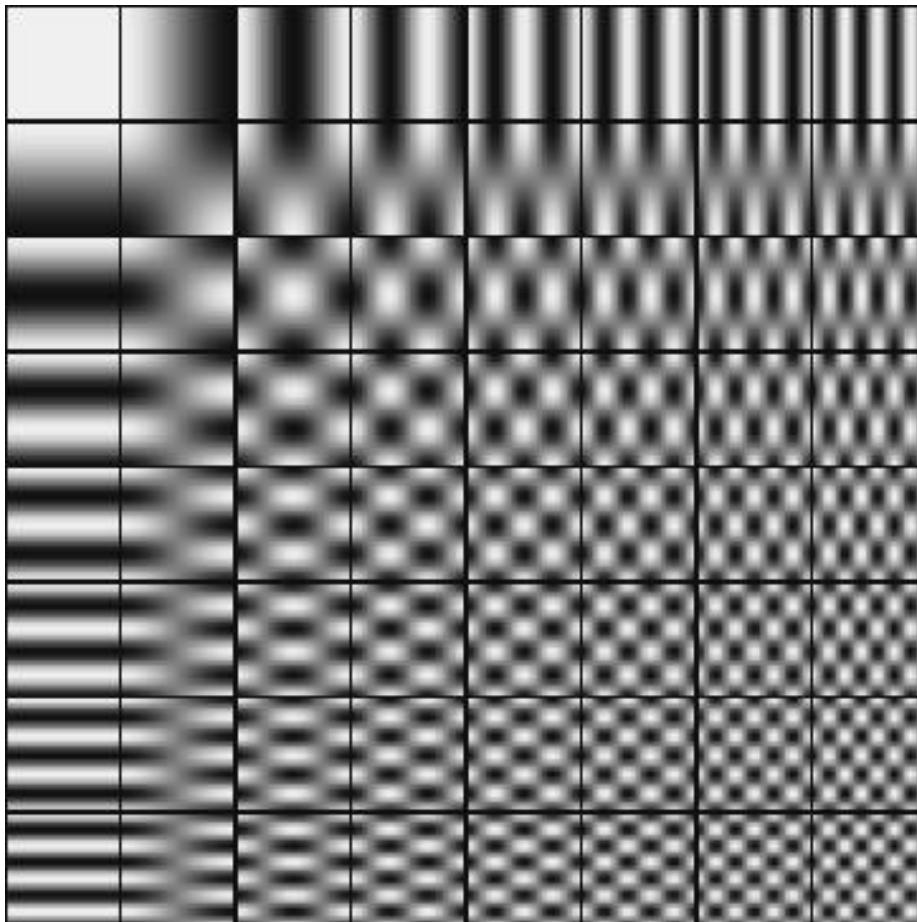


Рисунок 2.13 – Базисні функції двовимірного ДКП

На рис. 2.13 наведено візуальний образ 64 двовимірних функцій (кожний маленький квадрат) розмірності 8×8 , що утворюють базис ортогональних функцій двовимірного перетворення Фур'є. На наведеному рисунку ортогональні функції розташовано упорядковано у вигляді таблиці по вісім в горизонтальному і вертикальному напрямку. Двовимірні функції позначають двома індексами, що характеризують їх положення в упорядкованому базисі. Перший індекс визначає номер по вертикалі, а другий – по горизонталі. Амплітуду окремого відліку (характеризує яскравість, оптична густина або інший параметр окремого пікселя) в межах кожної ортогональної функції (на рис. 2.13) відтворено пропорційною градацією яскравості.

Розглянемо зв'язок між функціями одновимірного і двовимірного перетворення. В лівому верхньому куті знаходиться двовимірна функція з номером $(0; 0)$, що відтворює постійну складову сигнального розподілу. Розподіл яскравості пікселів цієї функції відповідають розподілу одновимірної функції з номером $n=0$, рис.2.12, в горизонтальному і вертикальному напрямках.

Розподіл яскравості пікселів двовимірної функції $(0; 1)$, що знаходиться праворуч від функції $(0; 0)$ відповідають розподілу одновимірної функції з номером $n=0$, рис.2.13, у вертикальному напрямку і горизонтальному функції з номером $n=1$ – в горизонтальному.

Аналогічно можна визначити зв'язок з одновимірними функціями для всіх інших двовимірних функцій

Застосуванням двовимірних косинусних функцій можна точно відтворити будь-який розподіл яскравості дискретизованого фрагменту зображення. Тобто для точного відтворення фрагменту зображення 8×8 пікселів нам треба використати 64 базисні двовимірні функції. Щоб отримати не спотворений фрагмент зображення треба попіксельно додати всі двовимірні функції з відповідними амплітудами (амплітуди відліків функцій визначаються коефіцієнтами перетворення).

Може виникнути питання: а в чому ж полягає вигащ від застосування ДКП? Якщо ми використовуємо відліки амплітуд, нам треба передати 64 значення

амплітуд, а якщо використовуємо ДКП – треба передати 64 значення коефіцієнтів. Виграш полягає в тому, що для опису фрагменту 8x8 елементів часто достатньо використати 1 або кілька функцій для його відтворення, тобто лиш є один або кілька коефіцієнтів перетворення будуть відмінні від нуля, а всі інші будуть дорівнювати нулю, як це наведено на рис.2.11. Для передавання нульових значень не треба використовувати багаторозрядне двійкове слово, а можна зробити це спеціальним чином (як це буде наведено далі). Тому застосування ДКП може скоротити сигнальний потік в кілька десятків разів.

2.3.7 Квантування коефіцієнтів ДКП

Для скорочення обсягу даних (стиснення) обчислені 12-розрядні значення коефіцієнтів ДКП замінюють зваженими та наближеними (повторно квантованими) значеннями з меншим числом розрядів. До квантованих у такий спосіб коефіцієнтів застосовують ентропійне кодування (кодування довжин серій).

Доцільно згадати, що зазначені вище перетворення здійснюють з дуже маленькими фрагментами зображення (8x8 пікселів). в телевізійному зображенні звичайної чіткості (SD) таких блоків укладається по горизонталі 90 (720 пікселів), а по вертикалі 72 (576 пікселів). Тобто в межах блоку мова йде про відтворення дуже маленьких фрагментів зображення. Їх розмір знаходиться на межі роздільної здатності зору. Для дуже дрібних деталей зір не розрізняє навіть доволі великі спотворення яскравості.

Властивості зору такі, що в кодері MPEG можна відкинути частину інформації з кодованого сигналу і це не призведе до помітної втрати якості відтвореного зображення. Коефіцієнти ДКП для високочастотних базисних функцій, можна передавати з меншою точністю, витрачаючи менше бітів.

Дослідження підтвердили, що у разі правильного вибору вагових матриць (матриць квантування)) сумарні спотворення практично не помітні за нормальної відстані спостереження (3–4 висоти екрану).

Найбільш точно (з округленням до 8-бітного формату) передають низькі частоти, а для підвищення точності передавання постійної складової (середнього значення всіх 64 пікселів блоку) застосовують спеціальний рекурсивний алгоритм, що ґрунтується на статистичному припущенні про близькість середніх значень суміжних блоків. Найвищі просторові частоти зазвичай квантують вдесятеро грубіше, ніж найнижчі.

Матрицю квантування сформовано у такий спосіб, щоб найменш точними (грубо квантованими) були високі діагональні частоти. Зрозуміло, внаслідок такого оброблення відносно малі отримані значення замінюють нульовими. З наведеного рис. 2.14 числового прикладу видно, що нулі зосереджено області високих і діагональних частот. Можливе застосування стандартної матриці квантування (за початковим налаштуванням) або матриці, запропонованої розробником кодера, що надходить безпосередньо з потоком стислих даних як службова інформація. Другий варіант зазвичай дає не тільки деяке (невелике) додаткове поліпшення коефіцієнта стиснення, але може істотно зменшити спотворення зображення.

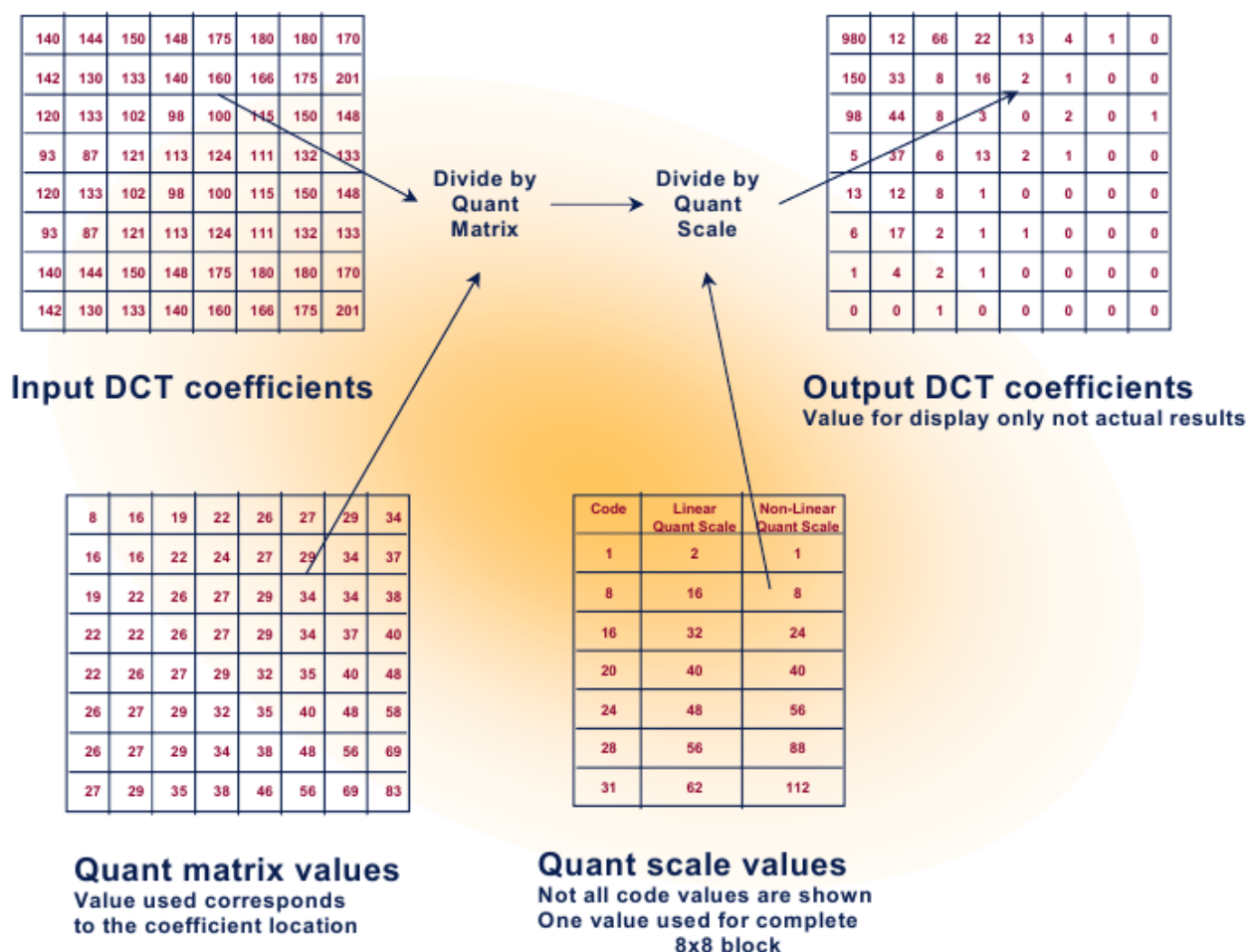


Рисунок 2.14 – Послідовність процедур квантування коефіцієнтів ДКП[]

Подальше стиснення обсягу переданих даних забезпечують методом кодування довжин серій та кодом змінної довжини (код Хафмана) (RLC/VLC = Run–Length Coding/Variable Length Coding). Довгі послідовності нулів, а загалом і будь–які значення, що з’являються з великою ймовірністю замінюють заздалегідь призначеними короткими кодами. Комбінації, що найчастіше зустрічаються, замінюють найкоротшими кодами.

Додатковий вигреш отримують зигзагоподібним зчитуванням (скануванням) коефіцієнтів ДКП, що називається «Z упорядкуванням» (zigzag), рис.2.15.

Сканування починають з постійної складової, потім найнижча горизонтальна частота, найнижча вертикальна частота, наступна вертикальна, потім найнижча діагональна частота, наступна горизонтальна тощо. Таким чином спочатку відбувається зчитування усіх ненульових коефіцієнтів, а потім зазвичай

йде довга серія безперервних нулів. Останню серію нулів блоку передають просто кодом кінця блоку, тобто, на передавання цієї серії витрачають мінімальну кількість бітів. Тим самим досягають значного стиснення загального обсягу даних.

Стандартом передбачено дві таблиці Z-упорядкування – «Стандартну» та «Альтернативну». Ефективність застосування цих таблиць є практично однаковою, хоча для черезрядкових відеоджерел застосування альтернативної таблиці може дати деяке (невелике) збільшення ступеня стиснення даних.

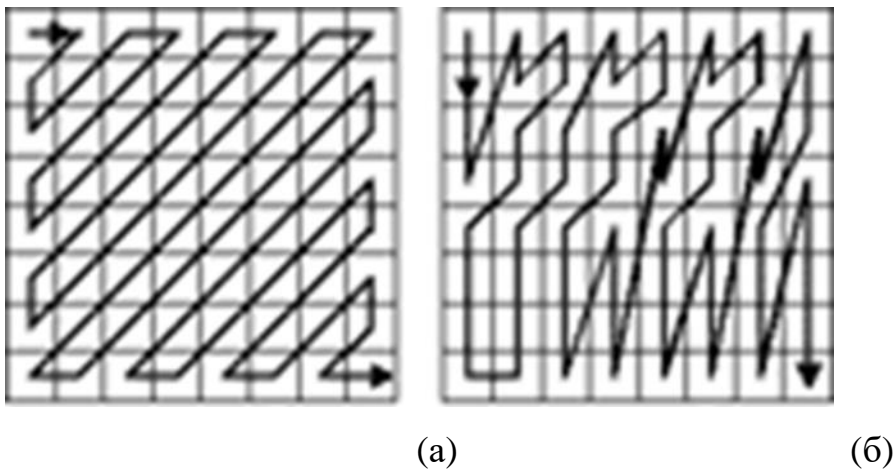


Рисунок 2.15 – Зигзагоподібна (а) та альтернативна (б) схема зчитування коефіцієнтів ДКП

Стосовно макроблоків подібних результатів досягають шляхом передавання коду керування «пропущеного» макроблока. У кадрах типу В «пропущені» макроблоки мають той самий вектор руху, що й попередній (базовий) макроблок. У кадрах типу Р «пропущеним» макроблокам приписують нульовий вектор руху.

Під час декодування (зворотному ДКП) грубо квантовані спектральні коефіцієнти призводять до спотворення зображення специфічними завадами. Зникають ізольовані дрібні деталі, з'являються паразитні візерунки як у межах блоків (blockiness = блоковість), так і усередині блоків («шахівниця»). В особливо складних випадках виникає відмінність яскравості або відмінність забарвлення сусідніх блоків – «мозаїчність» зображення, рис.2.16.

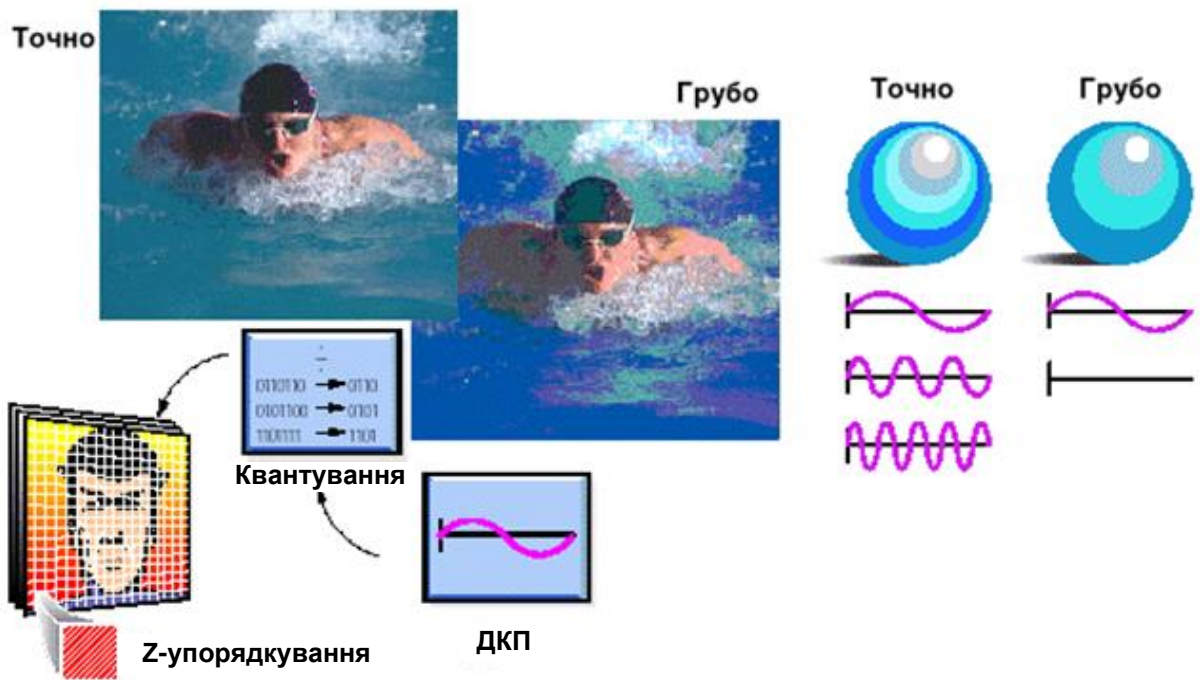


Рисунок 2.16 – Приклади спотворень, обумовлених наближеними значеннями коефіцієнтів ДКП

2.3.8 Колірні сигнали

Стандартом MPEG–2 передбачено три формати колірних сигналів, рис. 2.17:

- Формат 4:4:4 – матриці C_b , C_r та Y однакові по вертикалі та по горизонталі. Цей формат майже ніколи не застосовують.
- Формат 4:2:2 – горизонтальні розміри матриць C_b та C_r удвічі менші за розмір матриці Y , а вертикальні розміри однакові. Це так званий «професійний» формат, який застосовують під час виробництва програм.
- Формат 4:2:0 – як горизонтальні, так і вертикальні розміри матриць C_b і C_r вдвічі менші за розмір матриці Y . Саме цей формат найчастіше застосовують для мовлення ТВ програм, оскільки в порівнянні з форматом 4:2:2 дозволяє додатково знизити обсяг переданих даних приблизно на 25 %. Однак, застосування формату 4:2:0 не тільки знижує вертикальну чіткість кольору, але також практично ускладнює подальше студійне оброблення декодованого зображення з повторним стисненням.

Необхідно зазначити, що двозначне позначення «4:2:0» стосовно даного контексту означає і досить дивне взаємне розташування відліків Y і C_b , C_r . У разі формату 4:2:0 відліки колірних сигналів не тільки проріджено по вертикалі, а й

зміщено вниз на половину кроку відліків яскравості. Таке розташування відліків запропоновано для того, щоб полегшити вертикальне прикінцеве фільтрування в декодері під час відновлення повної вертикальної частоти кольорових відліків. Однак, для досягнення цієї мети довелось суттєво ускладнити процедуру попереднього фільтрування та кодер в цілому.

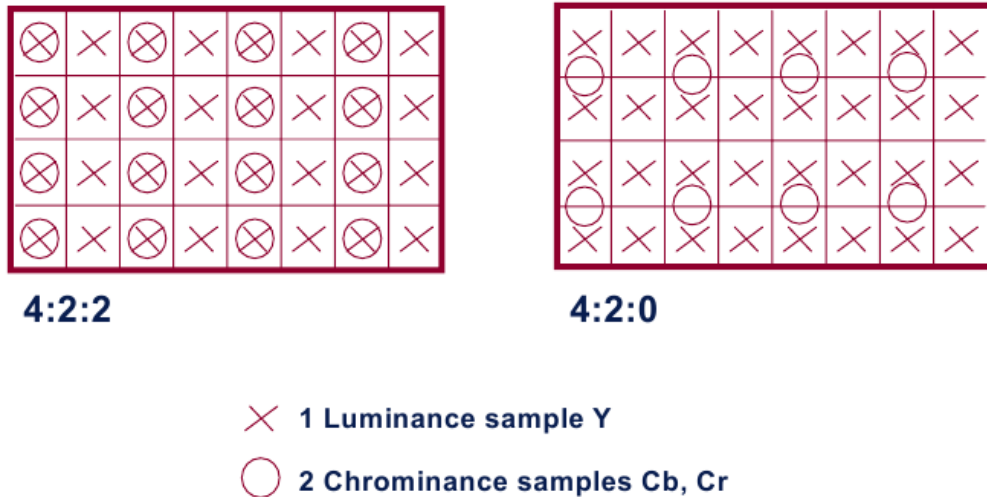


Рисунок 2.17 – Формати колірних сигналів[]

Чотири блоки яскравості поєднують з фіксованою кількістю просторово відповідних блоків колірності в межах одного макроблоку, рис. 2.18. Таким чином, загальна кількість блоків макроблоку залежить від вибраного формату колірних сигналів. Порядок оброблення та передавання блоків усередині макроблоку також залежить від формату колірних сигналів:

4:2:0 – макроблок складається із шести блоків. Містить чотири блоки яскравості та два блоки колірних сигналів (Cb та Cr) у такому порядку: Y1, Y2, Y3, Y4, Cb1, Cr1.

4:2:2 – макроблок складається з восьми блоків. Містить чотири блоки яскравості та чотири блоки колірних сигналів (по два блоки Cb та Cr) у такому порядку: Y1, Y2, Y3, Y4, Cb1, Cr1, Cb2, Cr2.

4:4:4 – макроблок складається з дванадцяти блоків. Містить чотири блоки яскравості та вісім блоків колірних сигналів (по чотири блоки Cb та Cr) у такому порядку: Y1, Y2, Y3, Y4, Cb1, Cr1, Cb2, Cr2, Cb3, Cr3, Cb4, Cr4.

	Y		Cb		Cr	
4:2:0	0	1				
	2	3	4		5	
4:2:2	0	1	4		5	
	2	3	6		7	
4:4:4	0	1	4	8	5	9
	2	3	6	10	7	11

Рисунок 2.18 – Послідовність блоків сигналів яскравості та кольорів під час оброблення одного макроблоку зображення (цифрами позначено блоки 8x8 відліків)

Макроблоки розміром 16 x 16 пікселів об'єднують по горизонталі в слайси, що зазвичай охоплюють всю ширину зображення. Слайс (slice = смужка, часточка, скибочка) складається з послідовності макроблоків зліва направо (далі будемо вживати термін «смужка»). В межах смужки основні параметри стиснення однакові, оскільки шкали квантування не можна змінюватися в межах смужки.

Смужки не можуть перекриватись, але можуть мати проміжки – пропущені макроблоки. Яким чином заповнювати чи відновлювати пропущені макроблоки і як декодер має реагувати на випадання смужки – це стандартом не обумовлено. Це важливе питання свідомо залишено на розсуд розробників декодерів.

Смужки об'єднують у відеокадри (pictures). У разі черезрядкової розгортки кадри розділяють на поля (напівкадри) й можливе як спільне, так і роздільне кодування відеоданих суміжних полів. Внутрішня організація макроблоків різна під час кодування полів та кадрів. У разі кодування полів блоки яскравості групують по полях: верхні першого напівкадру, а нижні – з другого. Блоки кольірних сигналів завжди групують в межах кадру, за винятком форматів 4:2:2 та 4:4:4, коли їх можна групувати як в межах кадру, так в межах поля. Перемикування

поля/кадри використовують для вибору методу обчислення векторів руху. Прогнозування в межах поля позначають англійською Inter-field-coding, тоді як прогноз в межах кадру позначається Inter-frame coding.

2.3.9 Регулювання заповнення буфера

Кодування довжин серій призводить до нерівномірної швидкості кінцевого потоку бітів на виході кодера. Однак в каналі передавання зазвичай треба забезпечувати безперервний потік з постійною швидкістю. Для узгодження швидкостей потоків необхідно використовувати проміжний буфер – як у кодері, так і в декодері, рис. 2.19. Переповнення або спустошення буфера є не припустимим. Для керування заповненням буфера застосований зворотний зв'язок, що регулює надходження бітів від кодера стиснення. Зміну інтенсивності потоку регулюють як шляхом більш тонкого або грубого налаштування квантувача, що обробляє коефіцієнти ДКП, так і іншими засобами. Для прийняття відповідних рішень щодо процесу керування стисненням в кодер вбудовано описану у стандарті «модель еталонного декодера». Декодер, зі свого боку, просто виконує командам кодера.

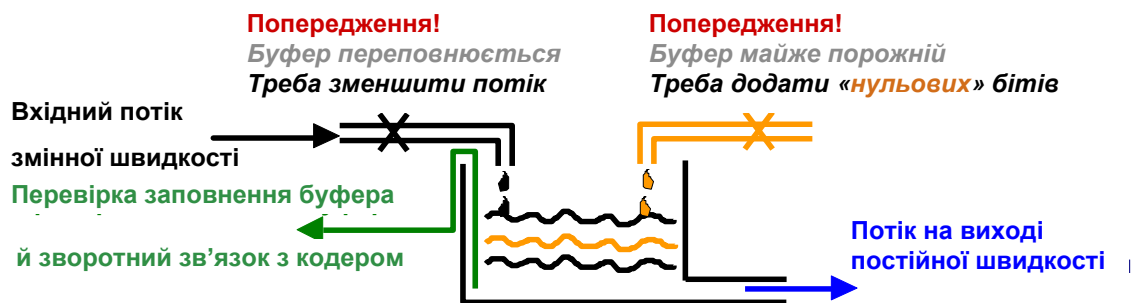


Рисунок 2.19 – Узагальнена схема керування проміжним буфером кодера

2.3.10 Основні елементи потоку стислих відеоданих

У процесі формування цифрового потоку стислих відеоданих блоки 8 x 8 елементів (сигналу яскравості або кольорних різницевих сигналів) перетворюють на коефіцієнти ДКП, кодують кодом довжин серій і об'єднують в макроблоки.

Кожен макроблок (16 елементів x 16 рядків) містить чотири блоки яскравості, пов'язані з ним блоки колірних сигналів (у форматі 4:2:0 це лише один блок Cr та один блок Cb), плюс заголовок, що містить вектори руху, опис режиму кодування та параметри квантування. Макроблоки об'єднують в смужки (слайси). Кожна смужка має стартовий код та адресу синхронізації. Якщо в межах даної смужки виявлено помилку передавання, декодер ігнорує смужку повністю і переходить до початку наступної смужки. Смужки утворюють відеокадри. Кожен відеокадр (picture) має стартовий код і заголовок, що містить інформацію про тип і структуру кадру, глобальні вектори руху, що передують композитному декодуванню тощо. Кадри групують в групи відеокадрів (GOP). Кожна група має стартовий код та заголовок, що містить інформацію про часовий код та редакторську інформацію. Групи GOP збирають у послідовності (sequences). Кожна послідовність має стартовий код і заголовок, що містить інформацію про швидкість потоку, розмір (роздільну здатність) зображення, формат кадру, частоту кадрів, коефіцієнт черезрядковості розгортки, формат колірних сигналів, та інші глобальні параметри кодування, зокрема – вказівку на використовуваний рівень і профіль кодування.

На рис. 2.20 наведено приклад форматування відеоданих у цифровому потоці.

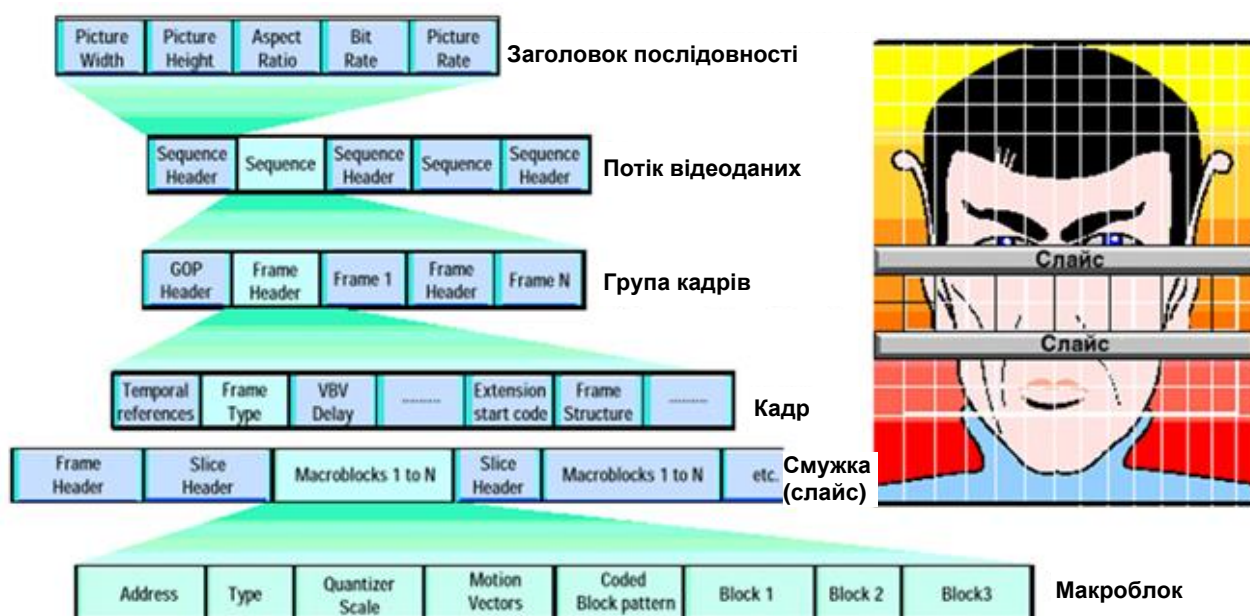


Рисунок 2.20 – Структура потоку стислих відеоданих

Слід зауважити, що деякі параметри форматування є обов'язковими елементами потоку (mandatory), інші – необов'язковими (optional). Крім того, є багато варіантів і режимів роботи. Наприклад, матрицю квантування можна передавати у вигляді набору з шістдесяти чотирьох 8-бітних коефіцієнтів, але за попереднім налаштуванням її можна використати повторно шляхом установавлення відповідного прапорця в заголовку послідовності. Єдина відмінність «послідовності» від «групи відеокадрів» полягає у наявності заголовка послідовності. Наприклад, послідовність може містити лише одну групу.

Інший приклад. В разі застосування режиму з постійною швидкістю цифрового потоку необхідно передавати спеціальну 16-бітну команду 'Керування буфером декодера – VBV_Delay (Video Buffer Verifier Delay = Рекомендована затримка буфера). Цією командою опосередковано задають початкове заповнення буфера, а число відліків протягом яких буфер має бути заповнений визначають за 90-кілогерцовим системним годинником. Умовний початок заповнення порожнього буфера відраховують від стартового коду кадру (Picture Start Code).

2.4 Стандарт H.264

2.4.1 Особливості стиснення відеоданих за рекомендацією H.264

Кодек H.264 містить практично ті ж елементи, що і кодеки MPEG1, MPEG2, MPEG4, H.261, H.263. В цілому стиснення відбувається за тими ж принципами, хоча у всіх основних елементах є якісні відмінності. Це дозволило значно підвищити якість кодування. За оцінками експертів, наприклад, більш ніж в два рази зростає ступінь стиснення за тієї ж якості відновлення початкового зображення.

Кодер (рис. 2.21) містить дві гілки оброблення відеопотоку:

- канал безпосереднього кодування (відзначений синім кольором), де оброблення даних згідно зі схемою здійснюють з лівого боку праворуч;
- канал реконструкції відеозображення (відзначений бузковим кольором), в

якому оброблення даних здійснюють з правого боку ліворуч. На рис. 2.22 наведено декодер. Оброблення даних роблять з лівого боку праворуч. У результаті порівняння рис. 2.21 і 2.22, можна побачити відповідність між блоками кодера і декодера.

На вхід кодера надходить кадр F_n . Оброблення кадру передбачено робити макроблоками, тобто фрагментами розміром 16×16 пікселів вхідного зображення. Для оброблення кожного макроблоку можна застосувати один з двох режимів: INTRA або INTER. У будь-якому режимі прогноз макроблоку P_n буде сформовано на підрунті відновленого кадру.

У режимі INTRA прогноз формують з вибірок поточного кадру n , попередньо закодованих і потім відновлених (F'_n на рис. 2.21 і 2.22). Для цього використовують вибірки кадру до відновлювального фільтра. У режимі INTER прогноз передбачено формувати з урахуванням змін, що мали місце в поточному кадрі порівняно з одним або кількома попередніми (або наступними). Кадри, які використовують для прогнозу, треба попередньо закодувати і відновити. У блоках формування INTRA– і INTER–прогнозу застосовують спосіб оброблення залежно від типу кодованого кадру. Ці способи детально описано в рекомендації, але критеріїв вибору не наведено. Завдання вибору вирішує розробник відеокодека самостійно.

Отриманий прогноз P_n попіксельно віднімають з поточного макроблоку. Таким чином обчислюють макроблок залишкових коефіцієнтів D_n . Цей макроблок надходить в перетворювач, де відбувається частотне перетворення залишкових коефіцієнтів. Необхідно відзначити, що рекомендацією передбачено використовувати квазіортогональне частотне перетворення. У такий спосіб вдалося значно зменшити обчислювальну складність алгоритму, але в процесі стиснення виникають додаткові спотворення. Частотні коефіцієнти квантують (масштабують), що призводить до стиснення відеоданих з втратами.

Отриманий набір перетворених і квантованих коефіцієнтів X є вхідним набором даних для зворотного каналу реконструкції даних. Надалі коефіцієнти пере– упорядковують в блоці спрямованого сканування. Коефіцієнти

упорядковуюють в лінійний масив в порядку зростання. Потім здійснюють ентропійне кодування упорядкованого масиву коефіцієнтів, що забезпечує стиснення даних без втрат. Закодовані коефіцієнти разом з додатковою інформацією, необхідною для правильного декодування макроблоку (режиму прогнозу, коефіцієнтів квантування тощо), утворюють стислий потік бітів (bitstream) абстрактного мережевого рівня (NAL). Цей потік можна передавати каналом зв'язку або записувати на будь-який носій для зберігання.

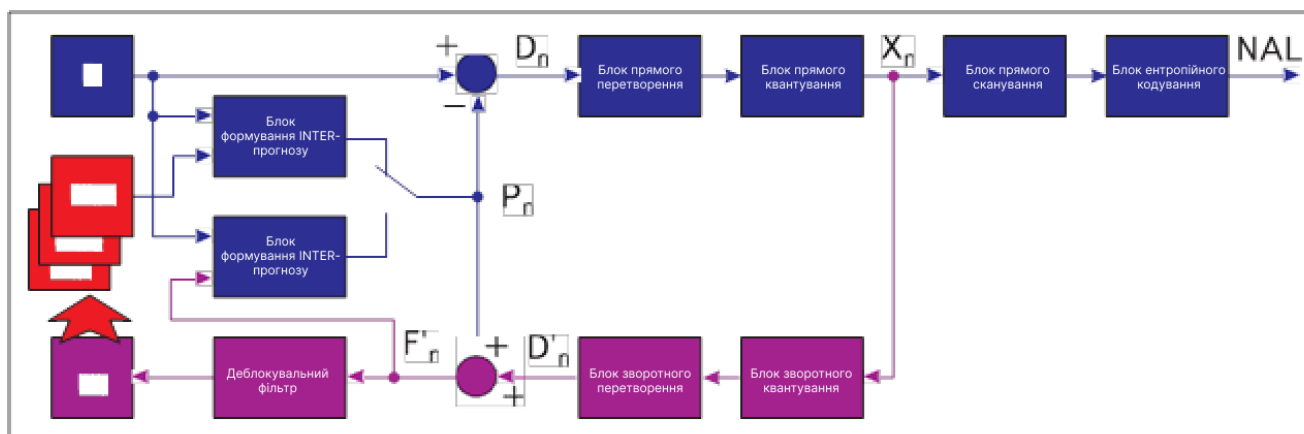


Рисунок 2.21 – Схема процесу кодування відеопотоку

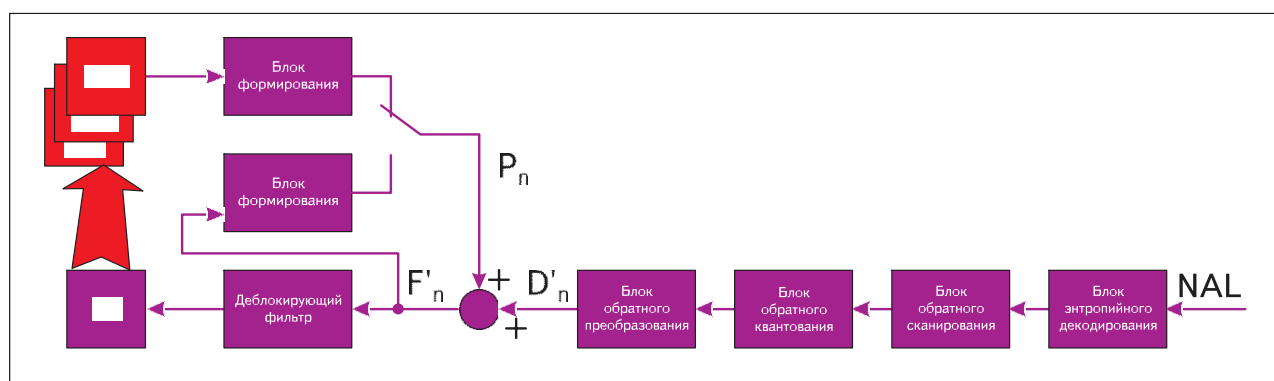


Рисунок 2.22 – Схема процесу декодування відеопотоку

У каналі реконструкції спочатку відбувається відновлення квантованих відліків, потім зворотне частотне перетворення. У підсумку отримують відновлені різницеві коефіцієнти D'_n . Їх додають до прогнозних значень P_n , і таким чином отримують відновлений кадр F'_n . Необхідно відзначити, що відновлений кадр не є ідентичним вихідному. До нього внесено спотворення, обумовлені

квантуванням і квазіортогональним частотним перетворенням. Саме такий кадр буде отримано на приймальній стороні, і тому саме його необхідно використовувати для формування прогнозу в режимі INTRA.

Під час оброблення кадру макроблоками виникають специфічні спотворення (блоковість), які проявляються в різких перепадах значень коефіцієнтів на межі макроблоків. Для їх зменшення застосовують відновлювальний фільтр. Отриманий після нього відновлений кадр використовують як опорний для формування прогнозу в режимі INTER. Відзначимо, що неможливо використовувати відновлювальний фільтр в режимі INTRA, оскільки на момент формування прогнозу в цьому режимі обробленими є тільки частина макроблоків, і повної інформації про межу між макроблоками нема.

Декодер отримує стислий сигнальний потік (bitstream) абстрактного мережевого рівня NAL (рис. 22.2). Структура декодера зрозуміла з рис. 2.22. В декодері передбачено зворотні процедури відносно до кодера. Варто зазначити, що в блоках формування INTRA– і INTER–прогнозів немає елементів, що потрібні для аналізу кадру. Інформація щодо конкретного режиму й способу його реалізації отримують з сигнального потоку (bitstream).

2.4.2 Формування внутрішньо–кадрового (INTRA) прогнозу

У режимі INTRA прогноз P формують на підставі попередньо закодованих і відновлених блоків. Для складової яскравості зображення прогноз P може бути сформований як для макроблоку розміром 16x16, так і окремо для всіх блоків розміром 4x4, з яких його складено. Всього визначено 9 додаткових режимів (способів) формування прогнозу (передбачення) для блоків яскравості розміром 4x4, 4 додаткових режими для макроблоків сигналів яскравості розміром 16x16 і 4 режими для блоків різницевих кольорних сигналів розміром 8x8.

Оброблення кадру в кодері здійснюють з правого краю ліворуч і згори вниз. Таким чином, формування прогнозу для поточного блоку (або макроблоку)

роблять із застосуванням вже оброблених блоків, розташованих згори і ліворуч від поточного (оброблюваного в даний момент). Відзначимо, що існує тільки один блок, для якого не використовують прогноз. Це блок, що розташований в лівому верхньому куті кадра. Крім того, прогнозування для блоків першого ряду оброблюваного кадру роблять тільки з використанням блоків, розташованих ліворуч, а прогноз для першого стовпця використовує тільки блоки розташовані над ними.

2.4.3 Режими формування прогнозу для блоків яскравості розміром 4x4

Відліки прогнозованого блоку (позначені на рис. 2.23 як a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p) формують із застосуванням вже оброблених вертикальних (I, J, K, L) і горизонтальних (A, B, C, D, E, F, G, H) відліків попередніх блоків (рис. 3).

Режими формування прогнозу:

Режим 0. Верхні відліки A, B, C, D екстраполують вертикально.

Режим 1. Ліві відліки I, J, K, L екстраполують горизонтально.

Режим 2. Всі відліки в прогнозі P є середнім з вибірок A ... D й I ... L.

Режим 3. Відлік інтерполують під кутом в 45° в напрямку між нижнім лівим і верхнім правим відліками.

Режим 4. Відлік інтерполують під кутом 45° вниз і праворуч.

Режим 5. Інтерполяцію відліків роблять під кутом приблизно $26,6^\circ$ ліворуч відносно до вертикального напрямку (відношення ширини до висоти дорівнює 1/2).

Режим 6. Інтерполяцію відліків роблять під кутом приблизно $26,6^\circ$ вниз відносно до горизонтального напрямку.

Режим 7. Інтерполяцію відліків роблять під кутом приблизно $26,6^\circ$ праворуч від вертикального напрямку.

Режим 8. Інтерполяцію відліків роблять під кутом приблизно $26,6^\circ$ вище горизонтального напрямку.

Стрілки на рис. 4 вказують напрямок передбачення в кожному режимі. У

режимах 3–8 відліки прогнозу формують з середньозважених вибірок А ... L. Наприклад, якщо обрано режим 4, то відлік прогнозу d (рис. 2.23) вираховують за формулою:

$$d = \text{round} \left(\frac{B}{4} + \frac{C}{2} + \frac{D}{4} \right)$$

(2.1)

	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				

Рисунок 2.23 – Маркування відліків для блоку прогнозних значень розміром

4 × 4

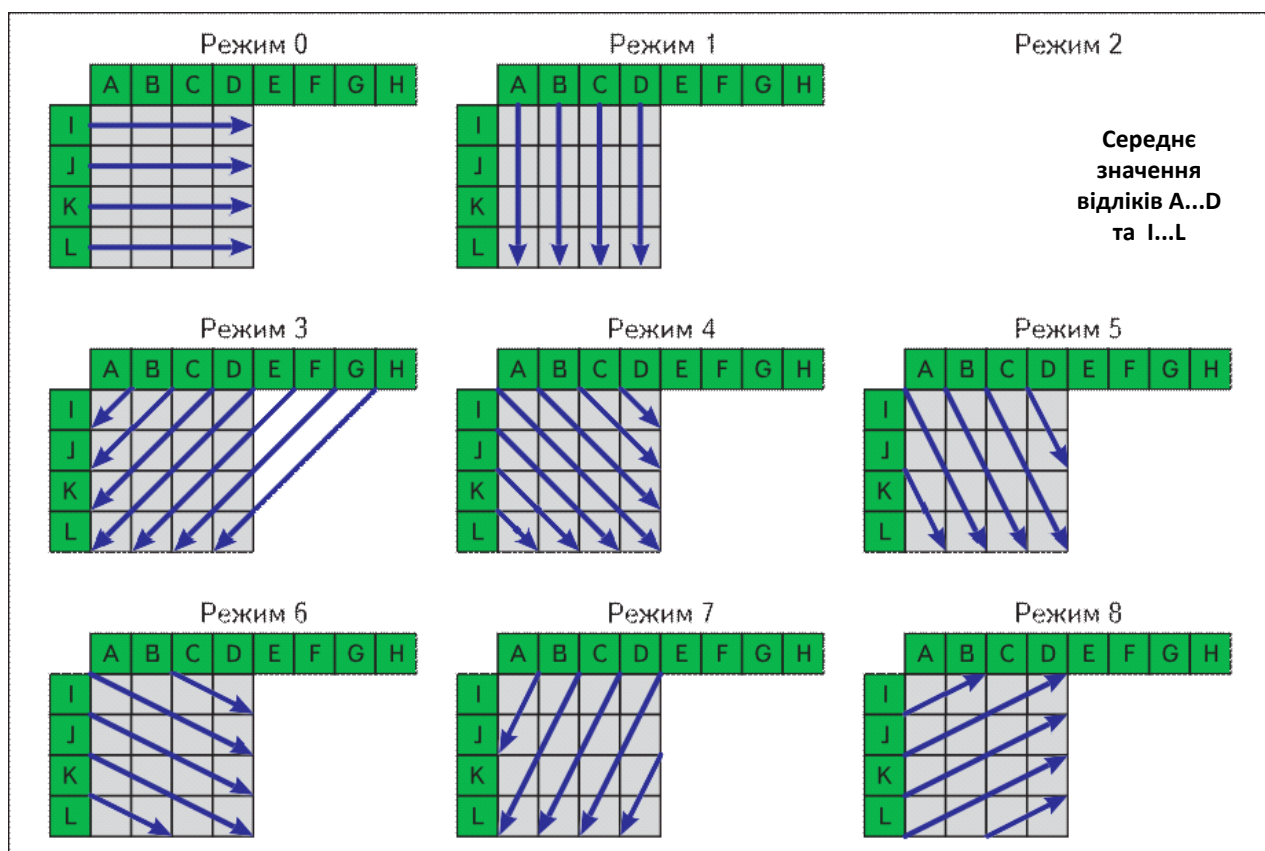


Рисунок 2.24 – Режими формування прогнозних значень для блоків

сигналу яскравості розміром 4×4

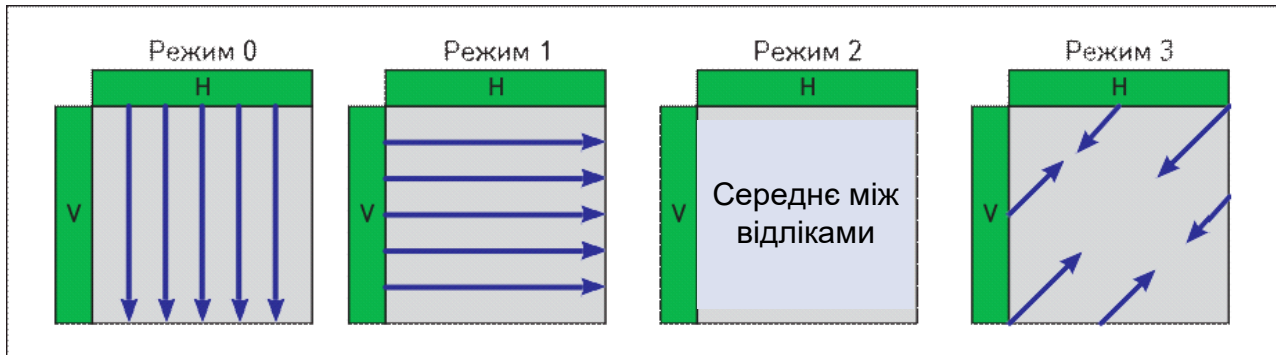


Рисунок 2.25 – Режим формування прогнозу INTRA (блоків розміром 16×16)

У кодері можна вибрати режим передбачення для кожного блоку таким чином, щоб мінімізувати відмінність між прогнозом P і блоком, для якого його визначили.

2.4.4 Режими формування прогнозу для блоків сигналу яскравості розміром 16×16

У кодері передбачено можливість кодувати не тільки блоки яскравості розміром 4×4 , але і розміром 16×16 . В рекомендації H.264 визначені чотири режими формування прогнозу для макроблоків (рис. 2.25).

- Режим 0. Екстраполяція верхніх вибірок H .
- Режим 1. Екстраполяція лівих вибірок V .
- Режим 2. Середнє між верхніми H і лівими вибірками V .
- Режим 3. Лінійне згладжування між верхніми H і лівими вибірками V .

Режими формування прогнозу для блоків різницевих колірних сигналів розміром 8×8 Для кожного з відліку блоку різницевих колірних сигналів розміром 8×8 прогноз формують із застосуванням різницевих колірних вибірок, розташованих вище і / або ліворуч, закодованих і відновлених. Чотири режими формування прогнозних значень для різницевих колірних складових аналогічні режимам для блоку сигналів яскравості розміром 16×16 , як це наведено у вступі.

Необхідно відзначити, якщо будь-який з блоків макроблоку сигналів яскравості кодується за внутрішньо-кадровим режимом (INTRA), тоді і обидва блоки різницевого кольорних сигналів також треба кодувати за цим режимом.

2.4.5 Формування внутрішньо-кадрового (INTER) прогнозу

У режимі внутрішньо-кадрового прогнозування відліки формують із застосуванням одного або кількох попередньо закодованих і відновлених кадрів шляхом вибору в них опорних блоків, зсунутих щодо кодованого блоку (прогноз з компенсацією руху). У кодеку стандарту H.264 використано той же принцип компенсації руху, що і в попередніх стандартах, починаючи з H.261. Важливими відмінностями від попередніх версій є можливість змінювати розмір блоку (від 16x16 до 4x4) і можливість пошуку векторів руху з меншим кроком (1/4 пікселя для яскравості компонент).

2.4.6 Розбиття макроблоку на блоки для формування прогнозу з компенсацією руху

У кодері, описуваному в рекомендації H.264, в режимі прогнозування з компенсацією руху передбачено можливість застосовувати блоків розміром від 16x16 до 4x4 для складової яскравості зображення в різних поєднаннях. Складову яскравості кожного макроблоку (розміром 16x16) можна розподілити чотирма способами, як це наведено на рис. 2.26: 16x16, 16x8, 8x16 або 8x8.

Кожна з областей меншого розміру є частиною макроблоку. Якщо обрано режим пошуку векторів руху для блоків 8x8, тоді кожен з чотирьох блоків 8x8 може бути розділений на блоки чотирма способами, як наведено на рис. 2.27, а саме: 8x8, 8x4, 4x8 або 4x4. Можливість ділити макроблок на блоки, а ті в свою чергу на субблоки, дозволяє отримати велику кількість комбінацій їх поєднань в межах кожного макроблоку. Цей метод розбиття макроблоків на блоки і субблоки в режимі компенсації руху отримав назву структурного дерева компенсації руху.

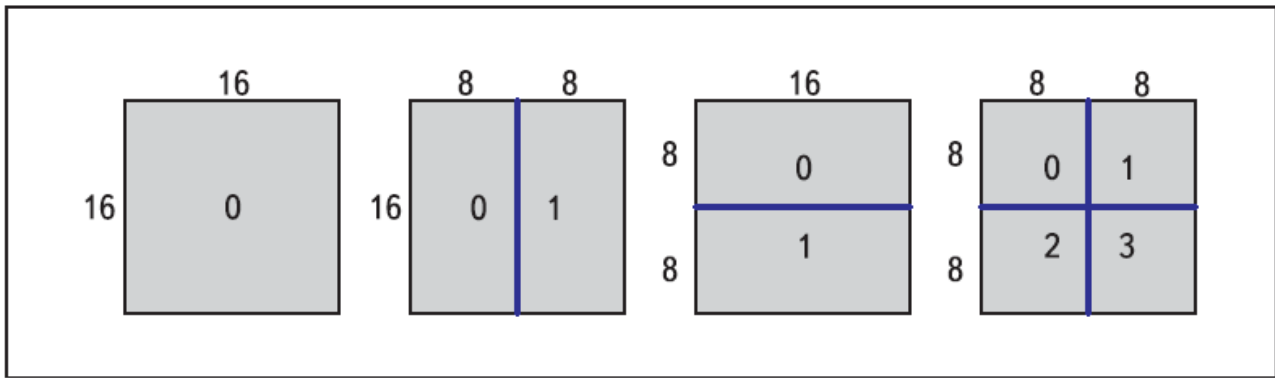


Рисунок 2.26 – Розділення макроблоку 16×16 на блоки 16×16 , 16×8 , 8×16 , 8×8

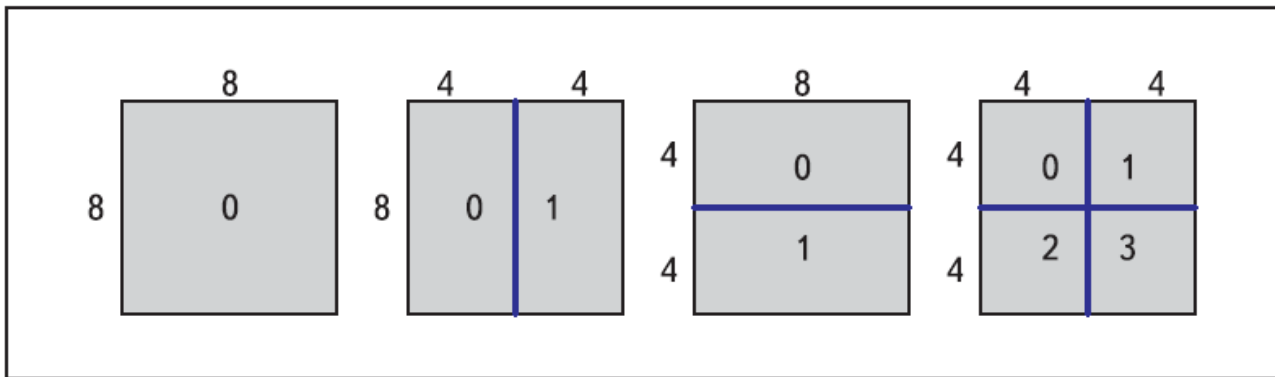


Рисунок 2.27 – Розділення блоку 8×8 на субблоки 8×8 , 8×4 , 4×8 , 4×4

Для кожного блоку і субблоку потрібен окремий вектор руху. Кожен вектор руху треба закодувати і передати до приймача; крім того, необхідно закодувати і передати конкретний вид структурного дерева для кожного макроблоку.

Вибір великого розміру блоків (наприклад, 16×16 , 16×8 , 8×16) означає, що кількість біт для передавання даних векторів руху і структури дерева буде мінімальним, проте, у такому разі, різницеві значення високочастотних складових (визначають детальність окремих пікселів блоку) можуть мати доволі великі амплітуди. У разі вибору маленького розміру блоків (наприклад, 8×4 , 4×4 тощо) можна отримати менші значення різницевих значень після компенсації руху, але потрібно буде використати більшу кількість бітів для кодування і передавання даних про вектори руху та структуру дерева. Тому, вибір розміру блоків має істотний вплив на ефективність стиснення. Загалом, великий розмір блоків доцільно застосовувати для однорідних частин кадру, що містять тільки

низькочастотні коефіцієнти перетворення, а маленький розмір блоків доцільно застосовувати для фрагментів кадру, що містять дрібні деталі і високочастотні коефіцієнти перетворення мають великі значення.

Лінійна роздільна здатність різницевих колірних складових макроблоку (C_r і C_b) дорівнює половині роздільної здатності складових яскравості. Кожен блок різницевих колірних вибірок виокремлюють таким же чином, як блоки вибірок яскравості, за винятком того, що вертикальні і горизонтальні розміри блоку в два рази менше (якщо блок вибірок яскравості має розмір 8×16 , то відповідний йому блок різницевих колірних вибірок має розмір 4×8 ; блоку вибірок яскравості 8×4 відповідає блок різницевих колірних вибірок 4×2 ; тощо). Горизонтальні і вертикальні складові кожного вектора руху (по одній на блок) розподіляють на два вектори, в разі застосування блоків різницевих колірних вибірок.

На рис. 8 наведено приклад розбиття кадру різницевих сигналів на блоки (до процедури компенсації руху). Кодер вибирає «кращий» розмір блоку для кожної частини кадру, тобто такий розмір, який мінімізує кодування різницевих відліків і векторів руху. На рис. 2.28 тонкими лініями наведено розподіл кадру на блоки. В областях кадру, де зміни незначні (різницеві відліки здаються сірими), вибрано розмір блоків 16×16 ; в областях з великими змінами (різницеві відліки здаються чорними або білими) вибрано менші розміри блоків.



Рисунок 2.28 – Приклад поділу кадру на блоки і субблоки

2.4.7 Дробові значення векторів руху

В режимі міжкадрового кодування (INTER) прогнозні значення для макроблоку формують із застосуванням значень відліків відповідного блоку того ж розміру в опорному кадрі. Зсув між цими двома областями (вектор руху) має мінімальну роздільну здатність, що дорівнює чверті відстані між точками (пікселями) в опорному зображенні (для складової сигналу яскравості). Опорних блоків сигналів яскравості і різницевих кольорних сигналів для не цілочисельних (дробових) векторів руху в опорному кадрі не існує, тому значення відліків для цих блоків необхідно обчислити із застосуванням значень сигналів близько розташованих пікселів найближчих пікселів.

На рис. 2.29 наведено приклад цілочисельного і дрібного векторів руху. Необхідно сформулювати прогноз для блоку розміром 4x4 (червоні точки на рис. 9а). Якщо горизонтальні і вертикальні компоненти вектора руху – цілі числа (рис. 9б),

то опорний блок в опорному кадрі існує (сині точки). Якщо один або обидва вектори руху – дробові числа (рис. 9в), прогноз (сині точки) формують значеннями, визначеними між суміжними пікселями в опорному кадрі (сірі точки).

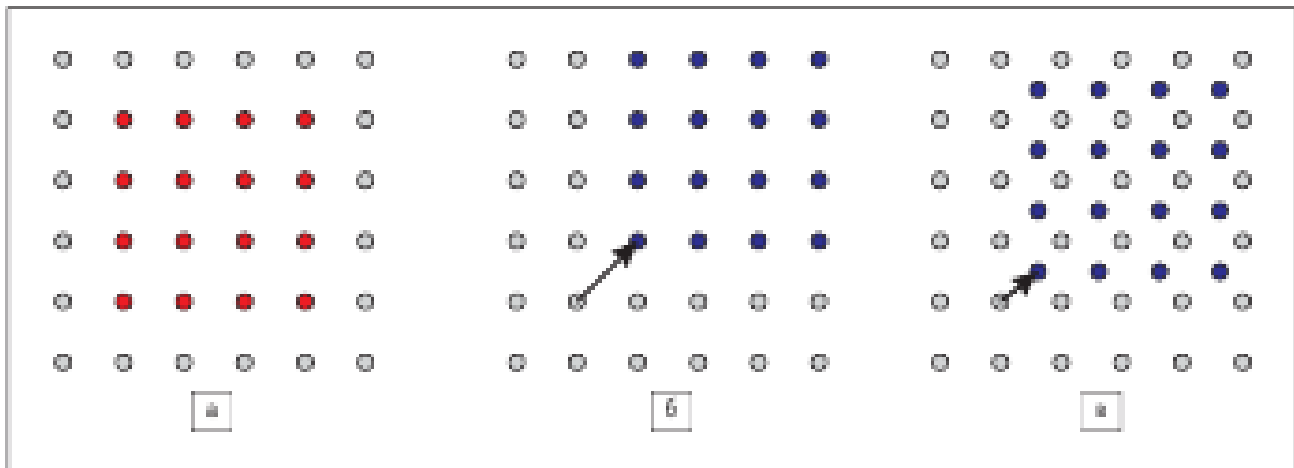


Рисунок 2.29 – Приклад цілочисельного і дробового векторів руху

Компенсація руху з четвертьпіксельною роздільною здатністю може забезпечувати значно кращу ефективність стиснення, ніж компенсація з цілочисловою роздільною здатністю, але в той же час збільшується складність процедури пошуку векторів руху.

Процедуру формування інтерпольованих вибірок наведено нижче. Для компонентів сигналу яскравості опорного зображення спочатку формують вибірки з напівпіксельною точністю (рис. 2.30). На рис. 2.30 вибірки в цілочислових позиціях виділено сірим кольором. Кожна вибірка в напівпіксельній позиції, яка є суміжною з двома цілочисловими вибірками (наприклад, вибірки *b*, *h*, *m*, *s* на рис. 2.30), інтерпольовані із застосуванням значень 6 пікселів в цілочислових позиціях із застосуванням КІХ-фільтра. Коефіцієнти фільтра мають такі значення: $1/32$, $-5/32$, $5/8$, $5/8$, $-5/32$, $1/32$.

Наприклад, напівпіксельну вибірку *b* розраховують із застосуванням шести горизонтальних цілочислових вибірок *E*, *F*, *G*, *H*, *I* і *J* за формулою:

$$b = \frac{\text{round}(E - 5F + 20G + 20H - 5I + J)}{32} \quad (2.2)$$

У такий же спосіб вибірку *h* інтерполюють на із застосуванням значень

вибірок A, C, G, M, R і T. Після формування напівпиксельних вибірок, суміжних з цілочисловими відліками, обчислюють інші напівпиксельні вибірки (наприклад, вибірки в позиції j на рис. 2.30) із застосуванням шести горизонтальних або вертикальних напівпиксельних вибірок, отриманих на першому кроці.

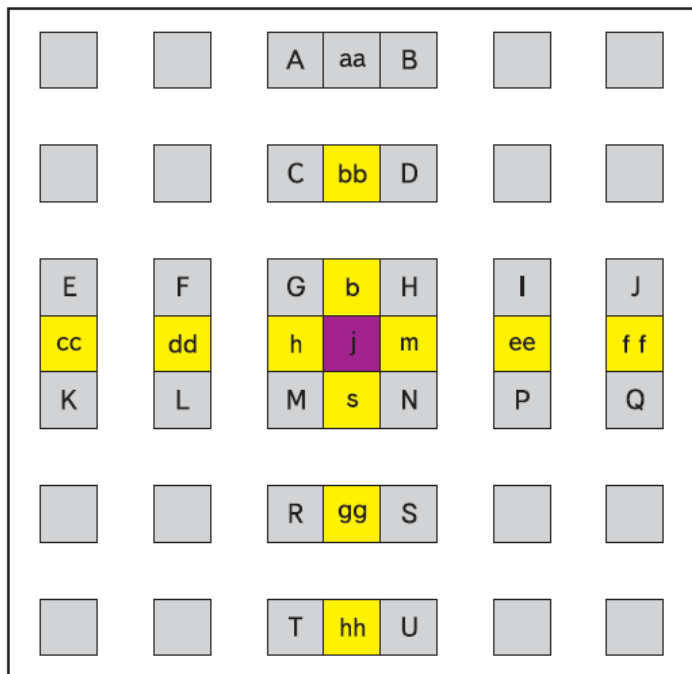


Рисунок 2.30 – Приклад розміщення напівпиксельних відліків в блок

Наприклад, відлік j сформовано із застосуванням значень вибірок cc, dd, h, m, ee і ff. (Зверніть увагу, що отримано однаковий результат у разі інтерполяції по горизонталі або по вертикалі.) Інтерполяційний КІХ-фільтр 6-го порядку є відносно складним (в порівнянні, наприклад, з білінійною інтерполяцією), але формує більш точний прогноз і, отже, забезпечує кращу ефективність компенсації руху.

Як тільки отримано всі вибірки в напівпиксельних позиціях, з їх застосуванням формують вибірки з чвертьпиксельною точністю методом лінійної інтерполяції (рис. 11). Вибірki в чвертьпиксельних позиціях формують із застосуванням двох горизонтальних або вертикальних суміжних напів- або цілочислових вибірок (наприклад, a, c, i, k і d, f, n, q на рис. 2.31) шляхом лінійної інтерполяції.

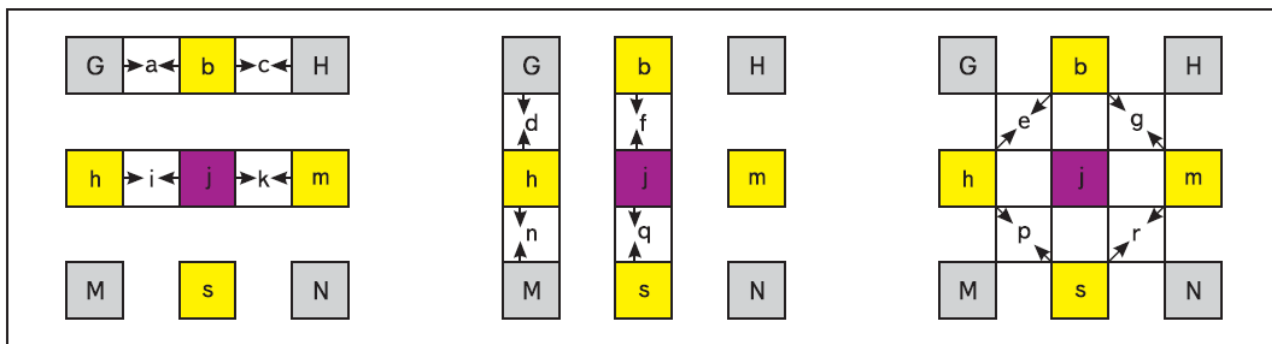


Рисунок 2.31 – Розміщення чвертьпиксельних вибірок в блоці сигналів яскравості

Наприклад:

$$a = \frac{\text{round}(G + b)}{2} \quad (2.3)$$

Решту чвертьпиксельних вибірок (e, g, p і r на рис. 2.31) лінійно інтерполюють між діагональною парою протилежних напівпиксельних вибірок. Наприклад, вибірку e інтерполюють із застосуванням вибірок b і h.

Якщо для компонентів сигналу яскравості вектор руху визначено з чвертьпиксельною точністю, тоді для компонентів різницевих колірних сигналів необхідно забезпечити 1/8-пиксельну точність визначення вектору руху (для формату кадру YUV 4: 2: 0). Інтерполяцію вибірок в 1/8-пиксельних інтервалах здійснюють із застосуванням цілочислових вибірок для кожного з блоків різницевих колірних сигналів. У цьому випадку використовують лінійну інтерполяцію для формування вибірок різницевих колірних сигналів з 1/8-пиксельної точністю (рис. 2.32).

У такому разі кожен вибірку формують як лінійну комбінацію сусідніх пікселів в цілочислових позиціях, наприклад, для вибірки a апроксимацію роблять із застосуванням цілочислових вибірок A, B, C і D (1).

Для прикладу можна розглянути випадок, наведений на рис. 2.32 ($dx = 2$ і $dy = 3$):

$$a = \frac{\text{round}(30A + 10B + 18C + 6D)}{64} \quad (2.4)$$

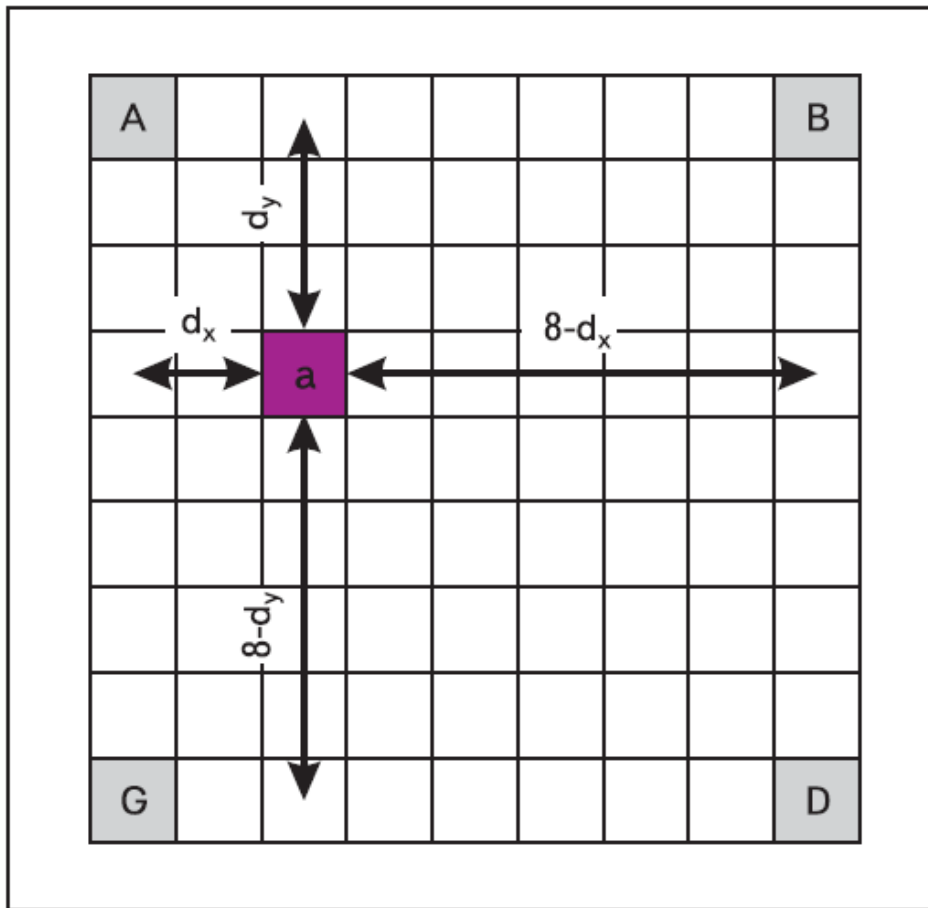


Рисунок 2.32 – Розміщення $1/8$ -піксельних вибірок в блоці сигналів яскравості

Вибірки кожного різницевого макроблоку перетворюють, сканують і квантують. У попередніх стандартах MPEG-1, MPEG-2, MPEG-4 і H.263 було передбачено використовувати дискретне косинусне перетворення (DCT) для масивів розміром 8×8 . В рекомендації H.264 визначено три види перетворення залежно від типу оброблюваних блоків:

- перетворення блоку розміром 4×4 нульових (DC) значень сигналу яскравості, отриманих з макроблоків сигналів розміром 16×6 , сформованих в режимі INTRA;
- перетворення блоку розміром 2×2 нульових (DC) різницевих колірних відліків, отриманих з блоків різницевих колірних будь-якого типу;
- перетворення для всіх інших блоків різницевих сигналів розміром 4×4 .

2.4.8 Сканування макроблоків

Сканування макроблоків є важливою частиною процесу стиснення відео за стандартом H.264. Макроблок – це блок відеокадру розміром 16x16 пікселів, який обробляється і передається в порядку, визначеному стандартом. Нижче розглянуті етапи сканування макроблоків, які використовуються при кодуванні H.264:

1. DC–коефіцієнти:

- У режимі INTRA для макроблоку розміром 16x16 спочатку передають блок DC–коефіцієнтів розміром 4x4.
- Ці DC–коефіцієнти вказують на зміну яскравості або інтенсивності зображення в межах макроблоку.

2. Блоки сигналу яскравості:

- Після передачі DC–коефіцієнтів передають інші блоки сигналу яскравості (0–15) в певному порядку.
- Ці блоки представляють різні частини макроблоку та містять інформацію про яскравість певних областей.

3. DC–коефіцієнти кольорових компонентів:

- Блоки 16 і 17 містять масиви розміром 2x2 DC–коефіцієнтів, отриманих з кольорових компонентів Cb і Cr відповідно.
- Ці DC–коефіцієнти вказують на зміни у колірних характеристиках зображення.

4. Кольорорізницеві блоки:

- Залишок кольорових компонент передається в залишених кольорорізницевих блоках (18 і 25).
- Ці блоки доповнюють інформацію про колір, що використовується для відтворення кольорового відтінку.

Цей порядок передачі даних визначається стандартом H.264 та допомагає забезпечити ефективне стиснення відео та оптимальну якість відтворення при декодуванні.

Висновки до розділу

H.264 та MPEG–2 є важливими у контексті стиснення відеоданих і мають значний вплив на відеоіндустрію.

H.264, відомий також як AVC (Advanced Video Coding), виявився значним кроком у вдосконаленні стиснення відео порівняно з попередніми стандартами, такими як MPEG–2. Однією з його ключових переваг є вища ефективність

стиснення, що дозволяє забезпечити високу якість відео при меншому обсязі даних. Це робить H.264 більш економічно ефективним для передачі відео через мережу, забезпечуючи високу якість при потребі в меншій пропускну здатності.

Другий стандарт, MPEG-2, був широко використаний у попередні роки, зокрема для трансляції телевізійних каналів та DVD. В порівнянні з H.264, він має меншу ефективність стиснення, що призводить до більшого обсягу файлу при збереженні високої якості. Також, MPEG-2 не має деяких сучасних функцій, які доступні у H.264, таких як більш вдосконалена обробка руху та покращений алгоритм прогнозування.

Порівнюючи два кодеки, слід відзначити, що H.264 став більш популярним завдяки своїй здатності забезпечити високу якість відео при менших розмірах файлів. Він є більш сучасним та ефективним стандартом, і використовується широко у сучасних технологіях, таких як стрімінгове відео, відеоконференції та інші.

Іншим важливим аспектом є застосування різноманітних технік компенсації руху та інтерполяції для прогнозування кадрів, що сприяє зменшенню обсягу даних при збереженні достатньої якості зображення. Техніки цілочислового та дробового векторного руху грають ключову роль у забезпеченні високоефективного кодування руху та зменшенні бітової швидкості.

H.264 виявився більш передовим і ефективним стандартом порівняно з MPEG-2, забезпечуючи високу якість відео при менших розмірах файлів та використовуючи сучасні техніки кодування.

3 ПРИНЦИПИ ОБ'ЄКТИВНОЇ ОЦІНКИ ОТРИМАНОГО ВІДЕОМАТЕРІАЛУ

3.1 Об'єктивна оцінка якості сприйняття відеоматеріалу

Об'єктивні методи вимірювання якості представляють у вигляді математичних алгоритмів, які описують процеси суб'єктивного оцінювання якості та засновані на критеріях і метриках, які можуть бути виміряні об'єктивно. Результати об'єктивних вимірювань повинні добре узгоджуватися з результатами суб'єктивного оцінювання для однієї і тієї ж відеопослідовності [1].

На практиці досить часто виникають ситуації, коли початкове і оброблене зображення здаються спостерігачеві ідентичними, тоді як об'єктивні методи для одних і тих же зображень дають дуже велику похибку. Враховуючи те, що оцінка якості людиною є вирішальною, подібна похибка під час об'єктивних вимірювань може бути не допустимою. Це обумовлює головну складність розробки об'єктивних методів [1].

Методи об'єктивного вимірювання якості зображень можна розділити на дві великі групи – методи, що не враховують особливості сприйняття зображень людиною, та методи, які враховують ці особливості. Методи першої групи більш прості для реалізації (пікове відношення сигнал/шум, середньоквадратична абсолютна різниця, середня різниця), проте методи другої групи дають найбільш відповідні суб'єктивним оцінкам результати [1].

3.2 Пікове відношення сигнал/шум

PSNR – Пікове відношення сигнал/шум. PSNR визначає рівень спотворень під час стиснення і включає підрахунок середньоквадратичної помилки (MSE). Діапазон прийнятих значень від 0 до 100, вимірюється в логарифмічній шкалі в децибелах, обчислюють за середньоквадратичним відхиленням початкового зображення від перетвореного відеокадру відносно квадрата найбільшого можливого значення пікселя $(2^n - 1)^2$, де n – кількість біт на відлік (піксель) кадру:

$$PSNR = 10 \cdot \lg \frac{(2^n - 1)^2 \cdot w \cdot h}{\sum_{i=1, j=1}^{wh} (x_{ij} - y_{ij})^2}, \quad (3.1)$$

де w – ширина відеокадру, h – висота відеокадру, x_{ij} , y_{ij} – значення пікселів початкового та обробленого зображень відповідно.

Що вище значення, то більше деталей залишилося у відеопослідовності після стиснення і, отже, вища якість.

PSNR – загальновідома проста метрика, яка не потребує складних обчислень, однак, різні дослідження показують невисоку кореляцію між значеннями метрики і фізіологічним людським сприйняттям.

PSNR фізично показує, в якому з випробовуваних варіантів залишилося більше деталей і менше шумів. Тому PSNR активно застосовується для реалізації завдань, пов'язаних із налаштуванням, оптимізацією та порівнянням енкодерів/транскодерів. Зокрема, якщо необхідно оперативно визначити, який з енкодерів/транскодерів забезпечує вищу якість кодування або за якого набору налаштувань енкодера/транскодера у відеопослідовності залишилося більше деталей.

3.3 Середнє пікове співвідношення сигнал/шум

APSNR (Average Peak Signal-to-Noise Ratio) є метрикою, яка використовується для об'єктивної оцінки якості відеоматеріалу. Вона базується на піковому співвідношенні сигнал/шум (PSNR), але враховує середнє значення PSNR для всіх блоків чи кадрів відео.

Співвідношення сигнал/шум (PSNR) вимірює, наскільки відбігає відтворена інформація (сигнал) від оригінальної інформації (шум), вираженої у децибелах. Високе значення PSNR вказує на високу якість відео, оскільки сигнал переважає шум.

APSNR розраховується шляхом усереднення значень PSNR для всіх блоків чи кадрів у відеопослідовності. Ця метрика дозволяє отримати загальний

показник якості відео на основі середніх значень PSNR.

Формула для APSNR:

$$APSNR = \frac{1}{N} \sum_{i=1}^N PSNR_i \quad (3.2)$$

де N – кількість блоків чи кадрів у відеопослідовності, а $PSNR_i$ – PSNR для кожного окремого блоку чи кадру.

APSNR є корисним інструментом при порівнянні якості відео, але важливо враховувати, що вона не завжди точно відображає сприйняття якості людиною, оскільки не враховує особливостей візуального сприйняття.

3.4 Середньоквадратична помилка

MSE (Mean Squared Error) – це метрика, яка використовується для оцінки якості узгодження між фактичними та прогнозованими значеннями в контексті різних сигналів чи даних. Вона розраховується як середнє квадратичне відхилення між кожним фактичним і прогнозованим значенням.

Формула MSE виглядає наступним чином:

$$APSNR = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (3.3)$$

N – кількість спостережень або точок даних,

Y_i – фактичне значення на позначці i ,

\hat{Y}_i – прогнозоване значення на позначці i .

MSE визначає середню величину квадратів відхилень між фактичними та прогнозованими значеннями. Чим менше значення MSE, тим краще відповідає прогноз фактичним даним.

Ця метрика широко використовується в областях, таких як обробка сигналів, машинне навчання та аналіз даних для кількісної оцінки точності моделей чи прогнозів.

3.5 Середньоквадратична абсолютна різниця

Mean Square Absolute Difference (MSAD) або середньоквадратична абсолютна різниця є метрикою, яка використовується для вимірювання різниці між двома наборами значень, зазвичай у контексті порівняння двох зображень чи сигналів.

У випадку обробки зображень, MSAD дозволяє оцінити, наскільки кожен піксель відрізняється між двома зображеннями. Формула MSAD подібна до формули середньоквадратичної помилки:

$$MSAD(X, Y) = \frac{\sum_{i=1, j=1}^{mn} |X_{i,j} - Y_{i,j}|}{mn}, \quad (3.4)$$

де m – кількість пікселів по горизонталі в відеокадрі, n – кількість пікселів по вертикалі у відеокадрі, $X_{i,j}$, $Y_{i,j}$ – значення пікселів початкового та обробленого зображень відповідно.

Менше значення MSAD вказує на більшу схожість між порівнюваними сигналами. У контексті зображень це може бути використано для оцінки якості відтворення чи для виявлення різниць між оригіналом і обробленим зображенням.

3.6 Середня різниця

Метрика вказує на усереднену абсолютну різницю між кольоровими компонентами відповідних точок порівнюваних зображень. Чим менше значення "середньої різниці," тим більше схожість між зображеннями. Такий підхід може бути корисним, наприклад, при тестуванні і налаштуванні кодеків або фільтрів для обробки зображень, коли важливо забезпечити якнайменше спотворень між оригінальним та обробленим зображенням.

$$d(X, Y) = \frac{\sum_{i=1, j=1}^{mn} (X_{i,j} - Y_{i,j})}{mn}. \quad (3.5)$$

MSAD та Delta можуть бути чутливими до невеликих різниць між

пікселями, і це може призводити до великої вартості метрик, навіть якщо ці різниці малі і невидимі для людського ока.

PSNR також може бути чутливим до невеликих змін, оскільки великі значення PSNR вказують на низькі рівні шуму, а високі рівні шуму можуть бути непомітними.

Усі ці метрики не завжди відображають те, як людина сприймає різницю між зображеннями. Наприклад, вони можуть недостатньо враховувати особливості сприйняття кольорів та контрасту.

В різних сценаріях MSAD, Delta та PSNR можуть давати різні результати, і їхні значення можуть бути важко інтерпретувати без врахування контексту конкретного застосування.

Хоча ці метрики допомагають оцінити якість зображень або сигналів, їхні недоліки підкреслюють важливість використання різноманітних методів оцінки якості, які враховують специфіку конкретного завдання та особливості відчуттів користувачів.

3.7 Індекс структурної схожості зображення

SSIM – метрика оцінки якості зображення за трьома критеріями: яскравість, контрастність і структура. Приймає значення від 0 до 1, при цьому чим вище значення, тим нижчі спотворення зображення і вища якість. Порівняно з PSNR, SSIM потребує більше обчислювальних ресурсів. SSIM – одна з перших успішних метрик, яка найточніше відповідає людському сприйняттю зображення, що підтверджується різними дослідженнями. Тому SSIM використовується при визначенні сприйманої якості, наприклад, при підтвердженні задовільної якості під час мовлення тощо. SSIM також застосовується на додаток до PSNR. Алгоритм обчислення індексу SSIM представлений на рис. 2.8.

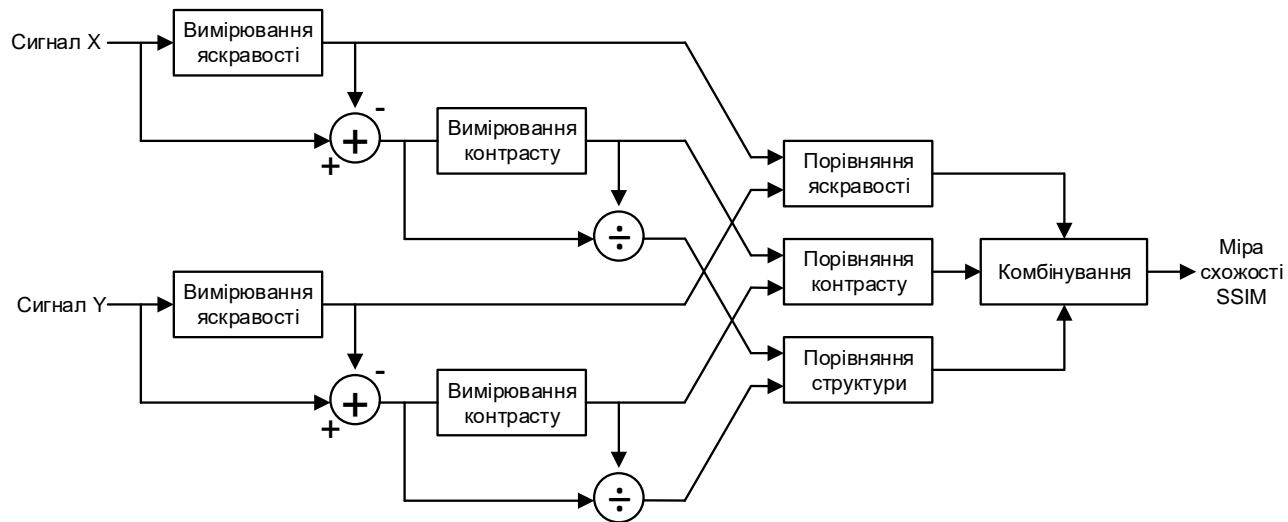


Рисунок 3.1 – Схема алгоритму обчислення індексу SSIM

Нехай $x = \{x_i | i = 1, 2 \dots N\}$ і $y = \{y_i | i = 1, 2 \dots N\}$ – еталонне і спотворене зображення відповідно. Тоді індекс структурної схожості зображення обчислюють згідно з виразом:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (3.6)$$

де α , β і γ – параметри, що визначають ступінь вагомості трьох складових індексу SSIM.

У виразі (2.18) міри схожості за яскравістю $l(x, y)$, за контрастом $c(x, y)$ і структурної схожості $s(x, y)$ представлені у вигляді:

$$l(x, y) = \frac{2\bar{x}\bar{y} + C_1}{\bar{x}^2 + \bar{y}^2 + C_1}, \quad (3.7)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (3.8)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (3.9)$$

де \bar{x} – середнє значення пікселів еталонного зображення як міра його яскравості:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (3.10)$$

\bar{y} – середнє значення пікселів спотвореного зображення, як міра його яскравості:

(3.11)

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i,$$

σ_x^2 – стандартне відхилення (девіація) значень пікселів еталонного зображення, як міра його контрасту:

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2, \quad (3.12)$$

σ_y^2 – стандартне відхилення (девіація) значень пікселів спотвореного зображення, як міра його контрасту:

$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2, \quad (3.13)$$

σ_{xy} – кореляція значень пікселів еталонного і спотвореного зображення, як міра структурної схожості двох зображень:

(3.14)

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}),$$

C_1, C_2, C_3 – деякі константи.

Якщо мають місце співвідношення $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, $C_3 = C_2/2$, де $K_1 \ll 1$, $K_2 \ll 1$, L – динамічний діапазон яскравості пікселів зображення (255 для 8-бітних зображень), та прийняти $\alpha = \beta = \gamma = 1$, тоді вираз, що описує індекс структурної схожості зображення, набуде вигляду:

$$SSIM(x, y) = \frac{(2 \cdot \bar{x} \cdot \bar{y} + C_1)(2 \cdot \sigma_{xy} + C_2)}{(\bar{x}^2 + \bar{y}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (3.15)$$

Визначення індексу SSIM відповідно до виразу (2.27) задовольняє трьом умовам:

- симетричності: $SSIM(x, y) = SSIM(y, x)$;
- обмеженості: $SSIM(x, y) \leq 1$;
- єдиного максимуму: $SSIM(x, y) = 1$ лише тоді, коли $x=y$.

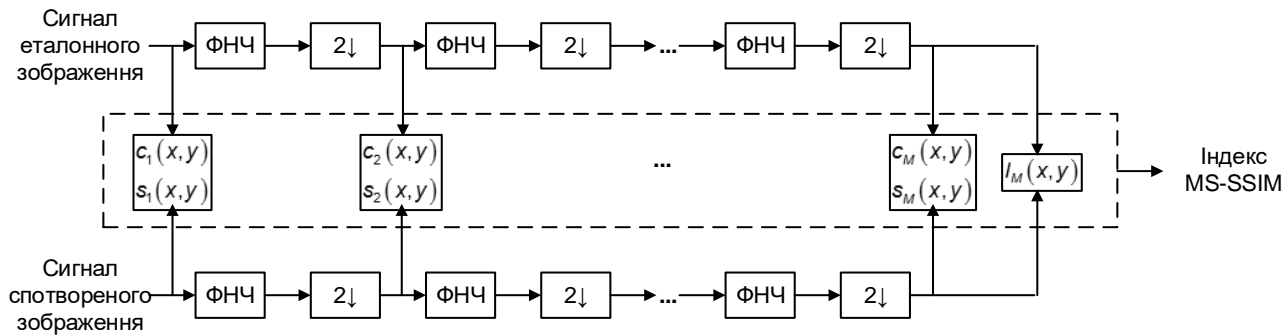


Рисунок 3.2 – Схема алгоритму обчислення індексу MS–SSIM

Сприйняття деталей зображення залежить від його роздільної здатності, відстані від площини зображення до спостерігача та сприйнятливості зору спостерігача. На практиці суб'єктивне оцінювання заданого зображення змінюється, коли змінюються ці фактори. Метод SSIM доцільно застосовувати лише за певних умов, оскільки він не враховує наведені вище фактори. Натомість метод багатомасштабного індексу структурної схожості (MS–SSIM) враховує деталі зображення за різної роздільної здатності [1].

Алгоритм роботи методу MS–SSIM наведено на рис. 2.9. Подаючи на вхід сигнали еталонного і спотвореного зображення, система ітеративно застосовує фільтр низьких частот і зменшує роздільну здатність відфільтрованого зображення вдвічі. Початковому зображенню присвоюють індекс зі значенням масштабного множника, що дорівнює 1, а зображенню після $M - 1$ рівнів фільтрації та пониження роздільної здатності, – індекс M . На j -й ітерації міру схожості за контрастом та міру структурної схожості розраховують та позначають як $c_j(x, y)$ і $s_j(x, y)$ відповідно. Міру схожості за яскравістю розраховують лише для індексу M і позначають як $l_M(x, y)$. Узагальнений вираз для обчислення індексу MS–SSIM має такий вигляд [1]:

$$MS-SSIM(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j}. \quad (3.16)$$

Експоненти α_M , β_j і γ_j аналогічно до виразу (2.18) використовують для

коригування ступеня вагомості різних складових. Визначення індексу MS–SSIM за формулою (2.28) задовольняє трьом умовам, визначеним вище. Метод SSIM є окремим випадком методу MS–SSIM, якщо покласти $M=1$.

3.8 Міра якості відео на основі дискретного косинусного перетворення

Алгоритм вимірювання якості відео (Video Quality Measurement – VQM) є методом для об'єктивної оцінки якості відеосигналів. Його основна ідея полягає в тому, що спостерігач під час оцінювання якості зображення зазвичай менше звертає увагу на дрібні деталі, а його увага концентрується на великих об'єктах. Такий підхід спроектований таким чином, щоб відображати відчуття споживача відносно якості відео.

Цей алгоритм може використовувати різні методи та підходи в залежності від конкретних вимог та особливостей вимірювання якості відео. Важливо також зазначити, що об'єктивні методи, такі як VQM, не завжди повністю відображають особисті відчуття споживача, тому їх використання слід розглядати як один з багатьох аспектів оцінки якості відео. Алгоритм обчислення міри VQM представлений на рис. 3.3.

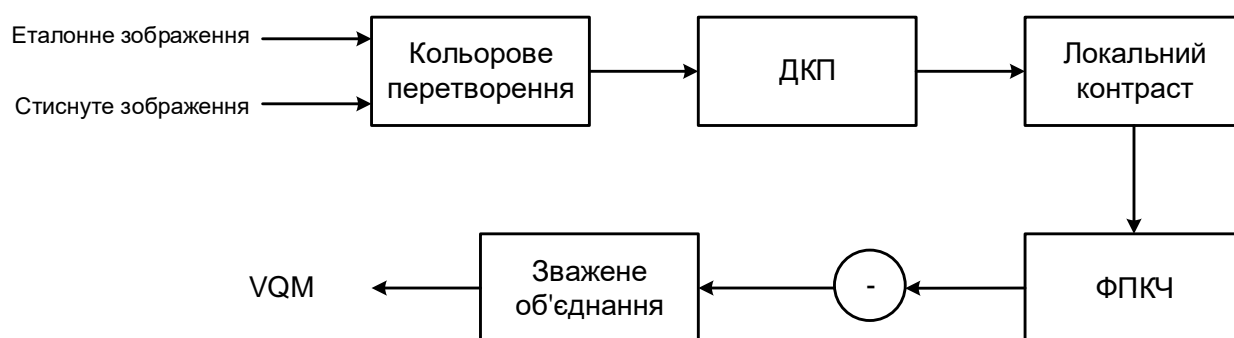


Рисунок 3.3 – Схема алгоритму обчислення міри VQM

Етапи алгоритму:

1. Розбиття відео на блоки: відеосигнал може бути розбитий на блоки, які служать для подальшого аналізу.
2. Виділення ключових об'єктів або областей: алгоритм може визначити

ключові об'єкти або області в кадрі, на які спостерігач зазвичай звертає більше уваги.

3. Алгоритм може використовувати різні методи для екстракції характеристик з кожного блоку чи ключової області, такі як контрастність, яскравість, колірність тощо. Результатом цього етапу є дві матриці локального контрасту (LC – Local Contrast) для блоків з початкового і спотвореного зображень відповідно:

$$LC(i, j) = \frac{DCT(i, j) \left(\frac{DC}{1024} \right)^{0.65}}{DC}, \quad (3.17)$$

де $DCT(i, j)$ – матриця розміром 8×8 , що є результатом дискретного косинусного перетворення початкового блоку, DC – це середня яскравість цього блоку перед перетворенням, що має нульову просторову частоту, тобто $DC = DCT(0,0)$.

4. На основі екстрагованих характеристик алгоритм може визначати значущість кожного блоку чи об'єкта з точки зору спостерігача.

5. Обчислюють функцію просторової контрастної чутливості ФПКЧ (Spatial Contrast Sensitivity Function – SCSF). Значущість кожного блоку або об'єкта може використовуватися для вагової оцінки якості відео, де більше уваги приділяється ключовим об'єктам.

6. Оцінки для всіх блоків чи областей можуть бути об'єднані для отримання фінального числового значення, яке представляє загальну якість відео.

$$VQM = 10 \left(\frac{1000 \cdot (\text{sum})}{(\text{width} \cdot \text{height})} + 5 \cdot \text{max} \right), \quad (3.18)$$

де sum – це сума усіх абсолютних відхилень, max – максимальне з усіх відхилень за усім кадром. Таким чином, оцінку якості зображення здійснюють з урахуванням максимального відхилення по усьому кадру, оскільки в алгоритмі зроблено припущення про те, що одне велике спотворення в одній частині зображення відверне увагу спостерігача від менших спотворень в інших частинах кадру.

3.9 Алгоритм визначення якості зображення на основі методу ледь помітних відмінностей Сарнова

Алгоритм PQR (Picture Quality Rating) ґрунтується на методі ледь помітних відмінностей (JND – Just-Noticeable Differences) Сарнова. Метод ЛПВ розроблений для оцінки того, наскільки невеликі зміни в зображенні можуть бути помічені спостерігачем. Це важливо для розуміння того, які впливи на якість зображення є важливими та помітними для людського ока.

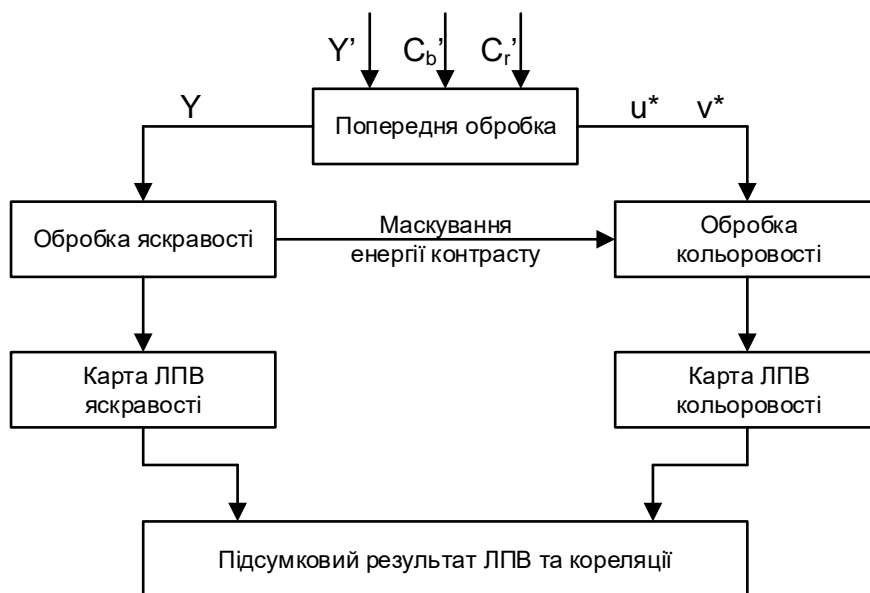


Рисунок 3.4 – Схема алгоритму обчислення PQR

Метод ЛПВ використовує концепцію порогового рівня відмінностей. Спостереження базується на тому, що споживачі можуть помічати тільки ті зміни в якості зображення, які перевищують певний поріг. Оцінка ЛПВ може бути використана для розробки алгоритмів, які моделюють та передбачають сприйнятливність споживачів до відмінностей у якості зображення.

Етапи алгоритму PQR можуть включати:

1. Вимір відмінностей між зображеннями:

Оцінка ЛПВ може включати вимір відмінностей між оригінальним та обробленим зображенням. Це може бути здійснене, наприклад, за допомогою

вимірювання різниці в яскравості, колірі, контрасті тощо.

2. Застосування порігового рівня:

За допомогою порогового рівня, визначеного методом ЛПВ, відфільтруються тільки ті відмінності, які спричинять помітні зміни для спостерігача.

3. Вагова оцінка:

Важливість кожної відмінності може бути визначена за допомогою вагових коефіцієнтів, які враховують їхню впливовість на загальну якість зображення.

4. Обчислення оцінки якості:

Оцінка якості PQR може бути обчислена за допомогою вагової суми помітних відмінностей.

Важливо враховувати, що конкретні етапи та параметри алгоритму PQR можуть варіюватися залежно від конкретної реалізації та використаного методу ЛПВ Сарнова.

3.10 Інші методи об'єктивного вимірювання якості зображень

Існують інші алгоритми обчислення якості відеопослідовностей, що ґрунтуються на об'єктивних метриках, проте вони мають вузьке призначення, оскільки дають можливість описати лише певне спотворення зображення.

Метрика мерехтіння між кадрами (BFM – Brightness Flicking Metric) є об'єктивною метрикою, яка призначена для вимірювання рівня мерехтіння між кадрами відеопослідовності. Мерехтіння може виникати внаслідок змін яскравості або інших параметрів зображення між кадрами, що може впливати на якість відтворення відео та сприйняття споживачем.

За результатами роботи методу будують графік покадрового значення метрики (рис. 2.12), де значення метрики – це абсолютна різниця між середньою яскравістю попереднього і поточного кадру. Підсумкове значення метрики є середнє арифметичне значення з усіх покадрових значень.

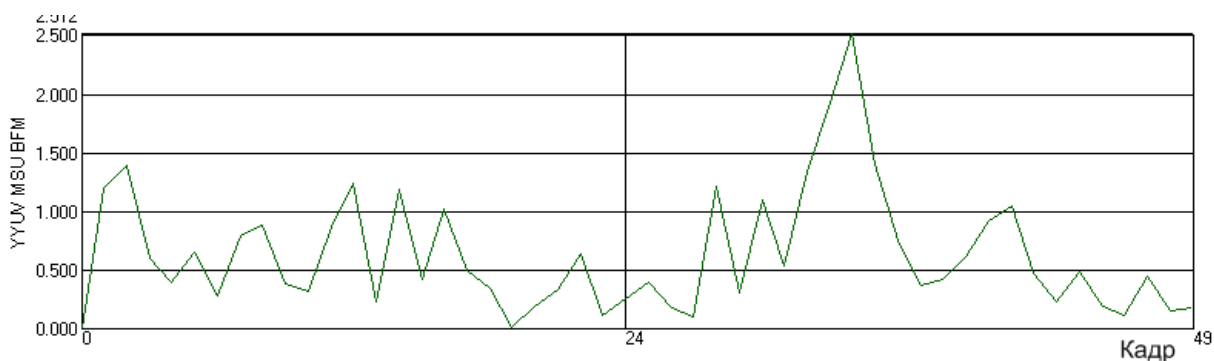


Рисунок 3.5 – Приклад графіка, що ілюструє метрику BFM

Метрика BFM може бути корисною для оцінки стійкості відеопослідовностей до мерехтіння та допомагати в розумінні його впливу на сприйняття якості відео.

Метрика пропущених кадрів (DFM – Drop Frame Metric) призначена для підрахунку кількості кадрів, які були пропущені кодером в послідовності відео. Вона використовується для оцінки ефективності кодування та виявлення проблем у відеопотоці, пов'язаних із пропуском кадрів.

Основні характеристики та етапи метрики DFM:

1. Визначення Міжкадрової Різниці:

Для кожної пари сусідніх кадрів визначається міжкадрова різниця в каналі яскравості, плюс 128. Це виконується для візуалізації змін яскравості пікселів між кадрами.

2. Побудова Графіка Покадрових Значень:

Результати аналізу міжкадрової різниці відображаються у вигляді графіка покадрових значень метрики, де ось X представляє собою час (по номерах кадрів), а ось Y представляє різницю в яскравості між кадрами.

3. Інтерпретація Графіка:

На графіку, 0 може вказувати на те, що кадр присутній, тоді як 1 – що кадр був пропущений. Таким чином, аналізуючи графік, можна визначити кількість пропущених кадрів у відеопослідовності.

4. Підсумкове Значення Метрики:

Як підсумкове значення метрики виводиться загальна кількість пропущених

кадрів від початку до кінця відеопослідовності.

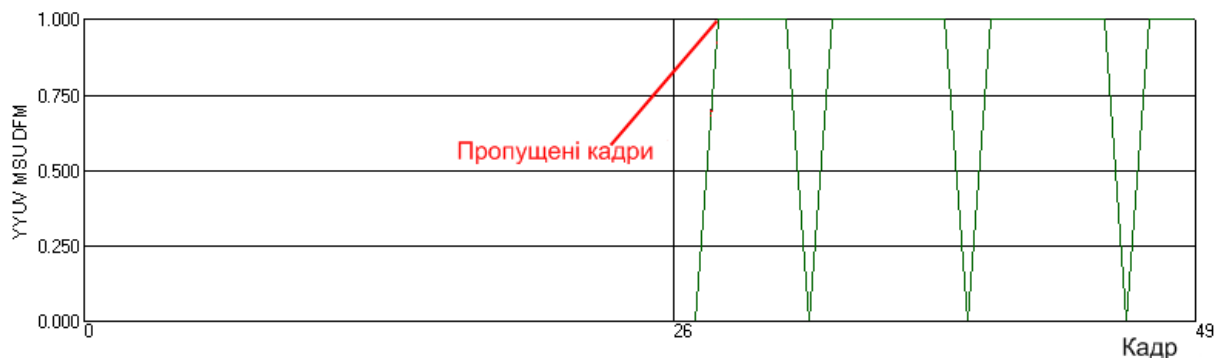


Рисунок 3.6 – Приклад графіка, що ілюструє метод DFM

Метрика оцінки рівня шуму (NE – Noise Estimation Metric) використовується для обчислення рівня шуму в кожному кадрі відеопослідовності. Метрика реалізує три різні алгоритми визначення рівня шуму: медіани абсолютного відхилення (MAD), блочної структури зображення (Block-Based) та просторово-часових гістограм (Spatio-Temporal Gradients). Нижче описано кожен з цих алгоритмів:

1. Алгоритм медіани абсолютного відхилення (MAD):
 - Для кожного кадру відеопослідовності застосовується вейвлетна декомпозиція Хаара.
 - Обчислюється медіана з абсолютних значень високочастотних коефіцієнтів (НН-компонент) після декомпозиції.
 - Підсумковим значенням метрики є нормована медіана.
2. Алгоритм блочної структури зображення (Block-Based):
 - Кожен кадр розбивається на блоки розміром 8x8.
 - Для кожного блоку обчислюється значення стандартного відхилення яскравості.
 - Значення стандартного відхилення сортуються, і обирається 30% блоків з найменшими значеннями.
 - Результатом метрики є нормоване середнє арифметичне 30% найменших значень.
3. Алгоритм просторово-часових гістограм (Spatio-Temporal Gradients):
 - Для кожного кадру використовується вейвлетна декомпозиція.

- Будуються просторова і часова гістограми.
- Первинна оцінка рівня шуму визначається значенням, коли часова або просторова гістограма досягає свого максимального значення.
- Рішення про використання гістограми визначається відхиленням гістограм від розподілу Релея.
- Ця оцінка корегується, використовуючи тест Колмогорова–Смірнова.
- Нормована скорегована оцінка є результатом метрики.

Метрика моментів зміни сцени (SCD – Scene Change Detector) призначена для автоматичного визначення моментів зміни сцени в послідовності кадрів. Вона реалізує чотири різні алгоритми побудови міри схожості сусідніх кадрів:

1. Піксельне порівняння кадрів (Pixel–level comparison):
 - Кожен піксель обох кадрів порівнюється на основі яскравості.
 - Різниця в яскравості визначає міру схожості між кадрами.
 - За результатами цього порівняння будується міра схожості.
2. Порівняння глобальних гістограм розподілу яскравості:
 - Глобальна гістограма розподілу яскравості обчислюється для кожного кадру.
 - Порівнюється схожість гістограм для визначення зміни сцени.
3. Поблочне порівняння гістограм розподілу яскравості (Block–Based Histogram):
 - Кадр розбивається на блоки.
 - Гістограма розподілу яскравості обчислюється для кожного блоку.
 - Порівнюється схожість гістограм блоків для визначення зміни сцени.
4. Побудову міри схожості на основі інформації про рух (Motion–Based):
 - Використовується інформація про рух між кадрами для визначення моментів зміни сцени.
 - Міра схожості визначається на основі змін у руховому векторі між сусідніми кадрами.

Під час візуалізації метрики відображають значення каналу яскравості, і яскравість першого кадру нової сцени штучно збільшується. Графік, побудований

за результатами роботи методу, позначає 1, якщо кадр є першим кадром у сцені, та 0 в інших випадках. Середнє значення метрики для послідовності дорівнює кількості знайдених змін сцени. Це допомагає в автоматичному визначенні та візуалізації моментів зміни сцени відеопослідовності.



Рисунок 3.7 – Приклад графіка SCD

NQI (Non-reference Quality Index) або HBI (Індекс якості без посилань) – це об’єктивна метрика, яка використовується для оцінки якості відео, і вона не потребує наявності оригінального (еталонного) відео для порівняння. Даний підхід особливо корисний в ситуаціях, коли оригінал не доступний або його важко отримати.

NQI враховує різні фактори, такі як артефакти стиснення, різкість зображення, колірна точність та інші параметри, щоб створити числовий показник, який відображає загальну якість відео. Основна перевага полягає в тому, що NQI може працювати без потреби у вихідних даних для порівняння.

Ця метрика використовується для автоматизованого аналізу та порівняння великої кількості відеоданих. Зазвичай вона застосовується в області відеостискання, телекомунікацій, відеонагляду та інших галузях, де важлива об’єктивна оцінка якості відеоматеріалу.

3.11 Класифікація та критерії оцінки об'єктивних методів вимірювання якості відеозображень

За обчислювальною складністю об'єктивні методи можна розділити на три групи. До першої групи можна віднести методи з *низькою обчислювальною складністю* (PSNR, MSAD, Delta, BFM, DFM, APSNR, MSE), які здійснюють прості математичні операції над яскравістю чи кольоровістю пікселів у кадрах.

До другої групи віднесено методи *середньої складності* (SSIM, MS-SSIM, NE, SCD, NQI), які ґрунтуються на простих математичних операціях над сигналами яскравості та кольоровості пікселів у кадрах для визначення кількох часткових показників з подальшим об'єднанням їх у єдиний показник якості. До третьої групи належать методи *високої обчислювальної складності* (VQM, PQR), що потребують виконання операцій обчислення ДКП або обробки великих масивів даних для побудови карт ЛПВ яскравості або кольоровості.

За ступенем відповідності суб'єктивним оцінкам якості виділяють методи з *низькою відповідністю* (PSNR, MSAD, Delta, BFM, DFM, NE, SCD, APSNR, MSE, NQI), які не враховують особливостей зору людини, та методи з *середньою та високою відповідністю* (SSIM, MS-SSIM, VQM, PQR), що навпаки враховують особливості зору людини.

За критерієм необхідності еталонного зображення можна виділити дві великі групи методів: методи, в яких використовують *еталонну відеопослідовність* для порівняння з досліджуваною відеопослідовністю (PSNR, MSAD, Delta, SSIM, MS-SSIM, VQM, PQR, MSE, PSNR), та методи, яким для роботи не потрібна еталонна відеопослідовність (BFM, DFM, NE, SCD, NQI).

Апаратна реалізація методів об'єктивного вимірювання якості відеозображень реалізована в обладнанні фірм OmniTek (метод PSNR) та Tektronix (метод PQR).

Методи BFM, DFM, NE, SCD, NQI не потребують для роботи еталонної відеопослідовності, що є безумовною перевагою цих методів. Проте вони є *вузькоспеціалізованими*, оскільки дають можливість оцінювати лише певні види

спотворень (наприклад, мерехтіння між кадрами, пропуск кадрів, зміну сцени тощо), які можуть бути несуттєвими для глядача. Натомість методи PSNR, MSAD, Delta, SSIM, MS-SSIM, VQM, PQR, APSNR, MSE дають можливість оцінити загальну якість відеозображення, що спостерігає глядач.

Ще одним важливим фактором, що може вплинути на вибір об'єктивного методу вимірювання якості, є його *доступність* у програмному середовищі. Ґрунтуючись на аналізі реалізації об'єктивних методів вимірювання якості у різних програмних пакетах, зокрема, у Elecard Video Quality Estimator, можна стверджувати, що переважна більшість методів, являється доступними для використання у програмному середовищі [1].

Висновки до розділу

Об'єктивні методи вимірювання якості можна класифікувати за обчислювальною складністю та ступенем відповідності суб'єктивним оцінкам. Перша група методів, яка включає PSNR, MSAD, Delta, BFM, DFM, NE, SCD, APSNR, MSE, NQI, характеризується низькою обчислювальною складністю і, в основному, і вони не враховують особливостей зору людини. Друга група, до якої відносяться SSIM, MS-SSIM, VQM, PQR має середню складність та враховує деякі аспекти зору людини. Третя група, яка включає VQM та PQR, є методами високої обчислювальної складності та враховує багато аспектів візуального сприйняття.

Важливим фактором при виборі методу вимірювання якості є його відповідність суб'єктивним оцінкам. Деякі методи забезпечують точніші результати, враховуючи більше аспектів сприйняття людиною, але вони можуть бути більш обчислювально витратними. Інші методи можуть бути менш точними, але при цьому мають низьку обчислювальну складність.

Методи, такі як BFM, DFM, NE, SCD, NQI, які не вимагають еталонної відеопослідовності, можуть бути корисними в специфічних випадках, але вони вузькоспеціалізовані та не забезпечують загальної оцінки якості.

Доступність методів у програмному середовищі є важливим аспектом при їх використанні. Більшість об'єктивних методів вимірювання якості доступні для використання у програмному середовищі, що робить їх доступними та зручними для досліджень та практичних застосувань.

4 ДОСЛІДЖЕННЯ КОДЕКІВ ЗА ДОПОМОГОЮ ОБ'ЄКТИВНИХ МЕТОДІВ ВИМІРЮВАННЯ ЯКОСТІ ВІДЕОЗОБРАЖЕНЬ

За умов гострої конкуренції за кожного користувача стримінгові платформи, оператори прагнуть забезпечити максимально високу якість відеоконтенту на всіх типах пристроїв і задіють із цією метою різні системи контролю якості відео. Багато таких систем як інструмент використовують об'єктивні метрики.

Відеопотоки проходять безліч етапів транскодування на шляху від правовласника до кінцевого глядача (рис. 4.1). Кожен етап стиснення призводить до втрати частини даних і зниження якості, а прагнення до зниження бітрейту може призвести до непотрібних артефактів стиснення. У зв'язку з цим виникає необхідність у використанні інструментів для контролю якості відео.

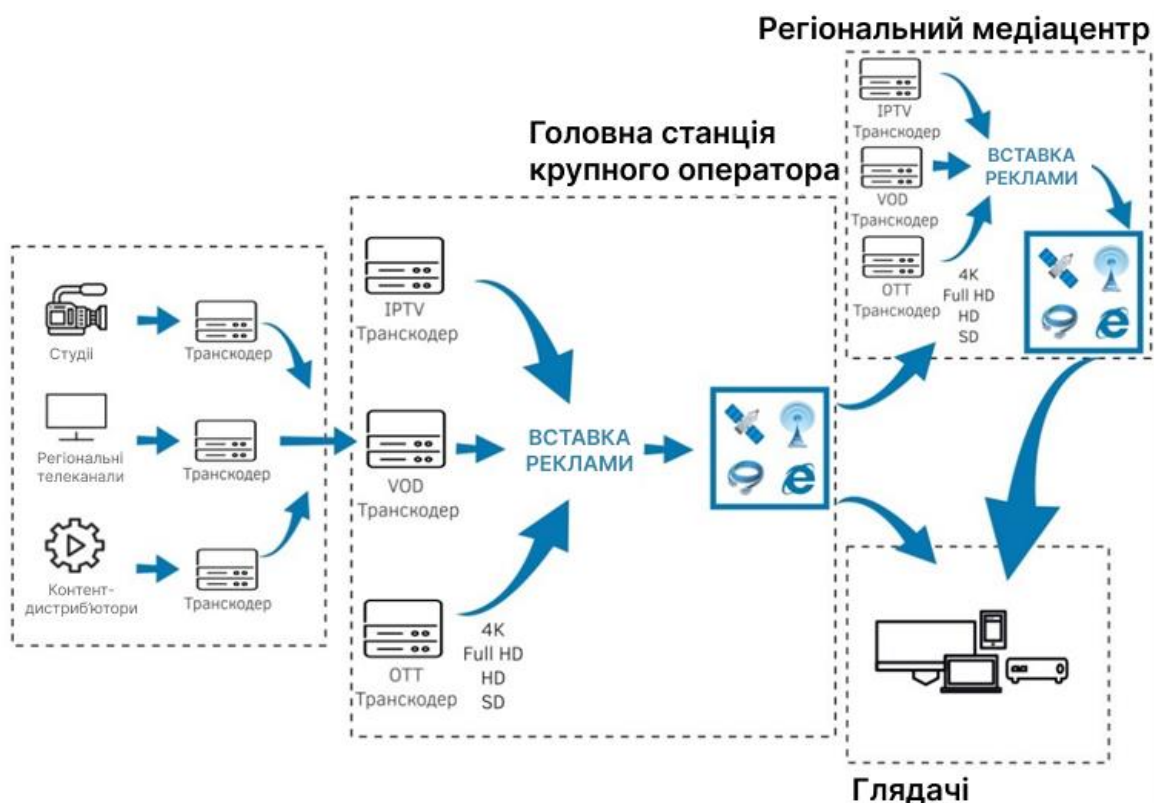


Рисунок 4.1– Шлях відеосигналу від джерела до глядача

Зазвичай метрики застосовують під час налаштування або порівняння енкодерів/транскoderів, а також під час контролю якості переданого сигналу.

Принцип роботи будь-якої об'єктивної метрики зводиться до підрахунку кількісної різниці між перекодованою та еталонною відеопослідовностями. Іншими словами, результуюче значення характеризує лише кількісну відмінність від оригіналу без оцінювання його якості у звичному розумінні.

4.1 Опис програмного забезпечення для вимірювання якості

Експериментальні дослідження з вимірювання якості відеозображень об'єктивними методами було проведено за допомогою програми Video Quality Estimator [2].

Elecard Video Quality Estimator – це потужний інструмент, розроблений для професіоналів і любителів у сфері стиснення відео. Elecard Video Quality Estimator дозволяє користувачеві розраховувати об'єктивні показники якості відео, такі як PSNR, APSNR, SSIM, DELTA, MSE, MSAD, VQM, NQI для порівняння двох потоків з вихідними даними. Зовнішній вигляд діалогового вікна програми наведено на рис. 4.2. Її інтерфейс інтуїтивно зрозумілий та містить покрокові інструкції з налаштування програми для запуску вимірювання якості відеопослідовностей за обраною метрикою:

1. На першому кроці необхідно вказати відеопослідовності, спотворені внаслідок стиснення (processed (compressed)).
2. На другому кроці необхідно вказати еталонну відеопослідовність (original file).
3. На третьому кроці обирають метрики, за допомогою якої буде здійснено об'єктивне вимірювання якості, та запускають процес оброблення (process). Зазначену метрику обчислюють як для кожного кадру, так і для всієї послідовності, отримуючи середнє значення метрики за усіма кадрами.

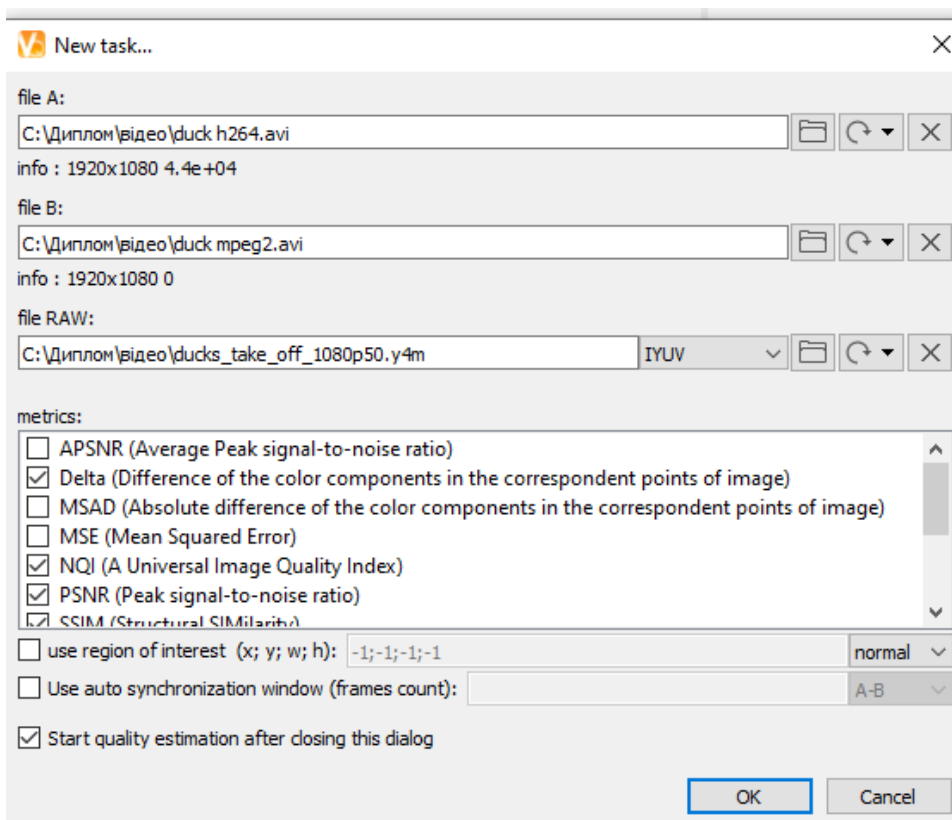


Рисунок 4.2 – Зовнішній вигляд діалогового вікна програми Elecard Video Quality Estimator

4.2 Постановка експерименту дослідження кодеків за допомогою оцінки якості зображення об'єктивним методом

Було вибрано 2 YUV відеопослідовності із глибиною кольору 8-біт, колірною субдискретизацією 4:2:0, роздільною здатністю 1920x1080 пікселів (рис.4.3).





а) б)

Рисунок 4.3 – Вибрані відеопослідовності:

а) Качки; б) Парк

Кожен екземпляр був стиснутий 2 кодеками. Було підготовлено 4 відеопослідовності у форматі AVC/H.264 та MPEG2, з глибиною кольору 8-біт, колірною субдискретизацією 4:2:0, роздільною здатністю 1920x1080 пікселів, фреймрейтом 50 кадрів на секунду, прогресивною розгорткою і тривалістю 10 с. Для кожного зразка було підраховано метрики: PSNR, SSIM, VQM, Delta, MSAD, MSE, NQI та APSNR за допомогою інструменту Elecard Video Quality Estimator. Середні значення метрик для кожної досліджуваної відеопослідовності, отримані в результаті експерименту, наведено у табл. 4.1.

Таблиця 4.1 – Середні значення об'єктивних метрик

Послідовність	Кодек	Середнє значення об'єктивної метрики							
		PSNR	MSAD	Delta	SSIM	VQM	NQI	APSNR	MSE
Качки	H.264	33,605	3,9699	0,23195	0,90024	1,9601	0,5187	33,605	28,662
	MPEG-2	29,437	6,1421	0,1065	0,77654	3,1772	0,31952	29,437	75,86
Парк	H.264	34,217	3,4234	0,20792	0,91875	1,9074	0,49198	34,217	25,087
	MPEG-2	29,069	5,722	0,15043	0,75842	3,6441	0,24467	29,069	82,848

Для прикладу нижче наведено графіки, які ілюструють покадрове значення відповідних метрик PSNR, SSIM та VQM для відеопослідовності «Качки» та «Парк» (рис. 4.5, 4.7, 4.9). Відеопослідовність, стиснута кодеком H.264 має зелений колір (А), стиснута кодеком MPEG–2 має зелений оранжевий колір (В). Повні результати покадрового вимірювання якості тестових відеопослідовностей за допомогою метрик PSNR, SSIM, VQM, Delta, MSAD, MSE, NQI та APSNR наведено у Додатку А.

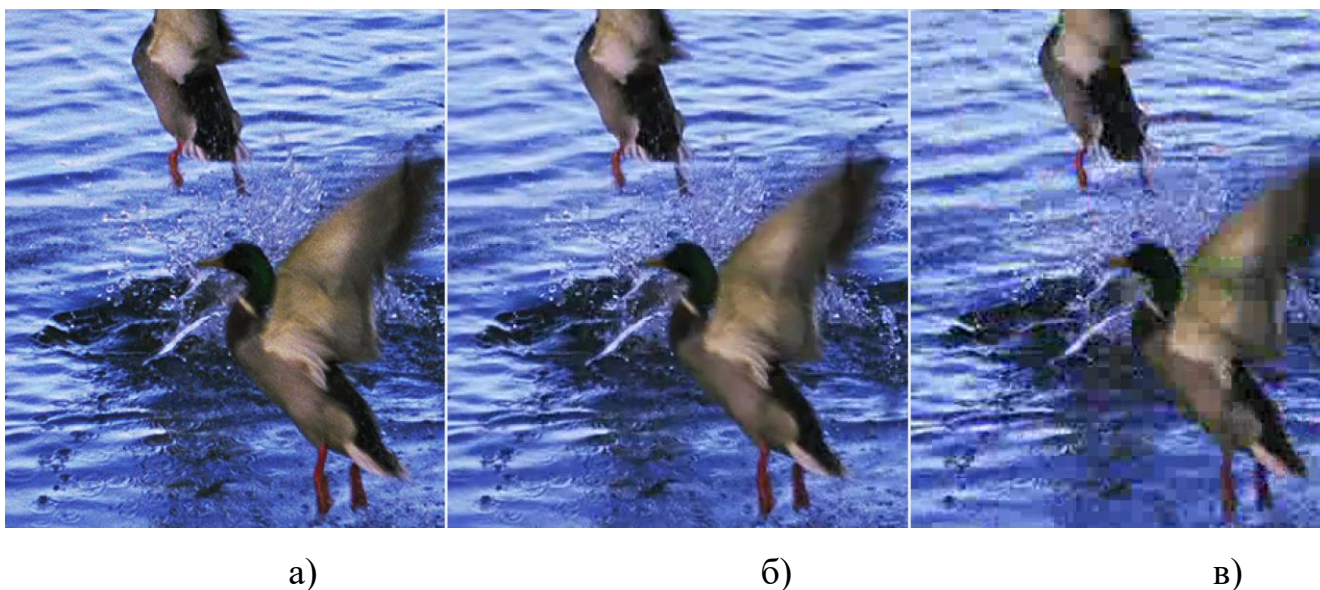


Рисунок 4.4 – Кадр відеопослідовності Качки

а) Оригінальна ВП; б) H.264; в) MPEG–2

На оригінальній відеопослідовності присутні шуми. Значення PSNR для H.264 свідчить про маленькі відмінності між оригіналом та стиснутим варіантом, кодек успішно приглушив шуми але виникла незначна втрата деталей та розмитість. Візуально відео, стиснуте за допомогою H.264, має високу чіткість та деталізацію, а відмінності від оригіналу майже непомітні.

Значення PSNR для MPEG–2 нижче, що вказує на більші відмінності між оригіналом та стиснутим варіантом. Це супроводжується помітними артефактами та втратами деталей. Суб'єктивно відео, стиснуте за допомогою MPEG–2, виглядає менш чітким та володіти меншою деталізацією порівняно з оригіналом.

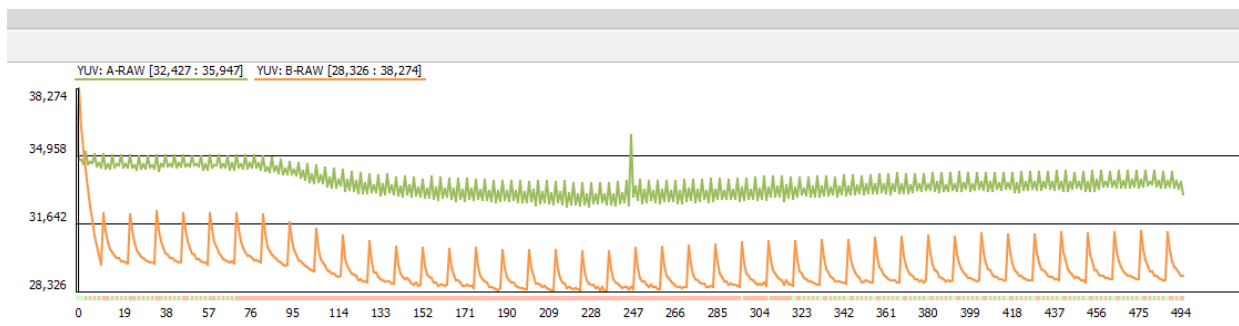


Рисунок 4.5 – Значення метрики PSNR для кадрів послідовності Качки



а)

б)

в)

Рисунок 4.6 – Кадр ВП Парк

а) Оригінальна ВП; б) H.264; в) MPEG–2

Високе значення SSIM для H.264 вказує на високу структурну подібність між оригіналом та стиснутим варіантом. Суб'єктивно це означає, що відео залишається структурно вірним оригіналу, і відмінності важко помітні для спостерігача. Структурна подібність сприяє відчуттю вірності відтворення та відсутності помітних артефактів, втрати деталей, чіткості контурів об'єктів або важкості визначення їхніх меж, зміни в деталях та текстурі поверхні об'єктів.

Низьке значення SSIM для MPEG–2 вказує на меншу структурну подібність. Суб'єктивно це означає, що відео має більше відмінностей у вигляді структурних артефактів чи втрати структурних елементів. На кадрі червона парасолька втратила чіткості контуру. Вцілому, спостерігається втрата чіткості контурів,

малих деталей, які можуть бути важливими для загального сприйняття зображення.

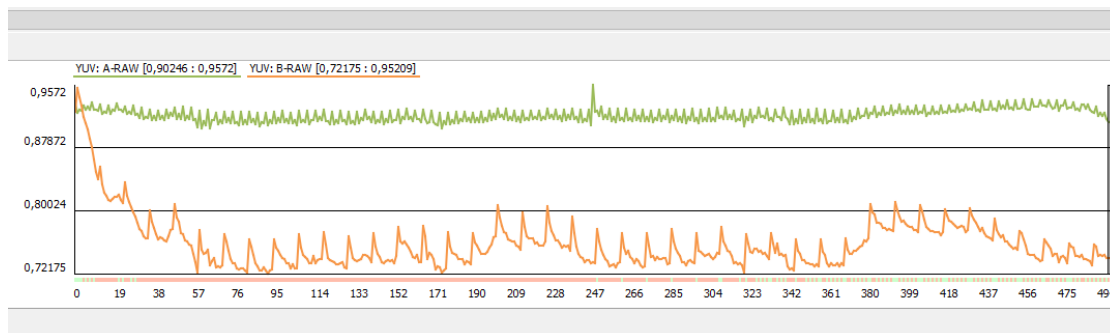
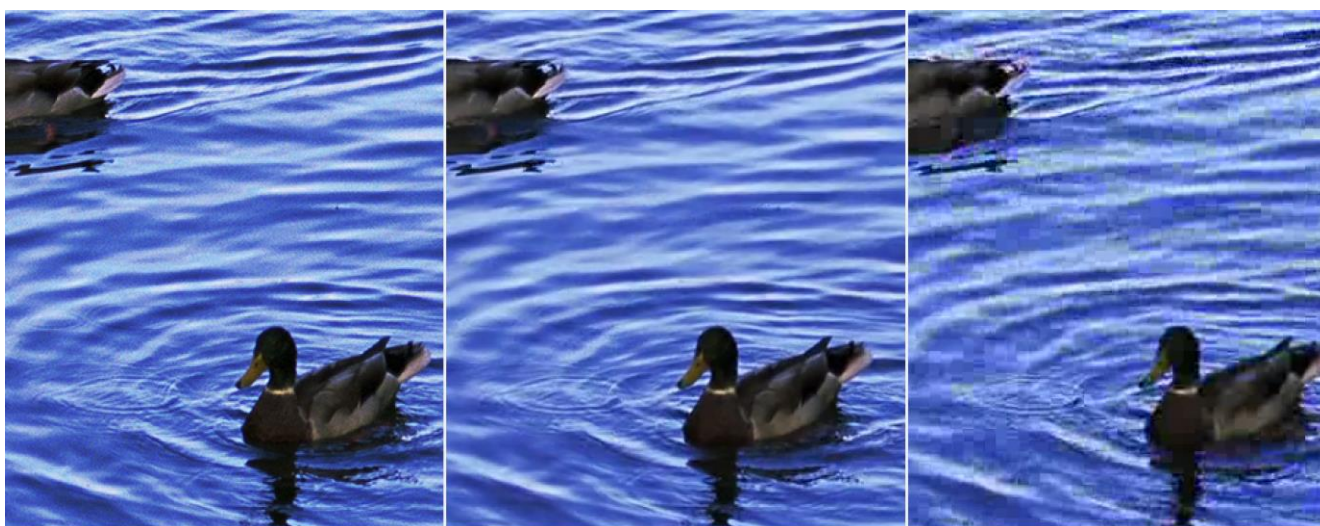


Рисунок 4.7 – Значення метрики SSIM для кадрів послідовності Парк



а)

б)

в)

Рисунок 4.8 – Кадр ВП Качки

а) Оригінальна ВП; б) H.264; в) MPEG-2

Низьке значення VQM для H.264 свідчить про високу якість відео з точки зору об'єктивної метрики. Суб'єктивно це може означати відсутність помітних артефактів, чіткість та стабільність відображення.

Високі значення VQM можуть вказувати на спотворення, розмиття, або артефакти на зображенні. Наприклад, це може включати блокові артефакти, аномалії в кольорі, розмиття контуру об'єктів, артефакти руху та інші аномалії, які можуть бути помітні при перегляді відео.

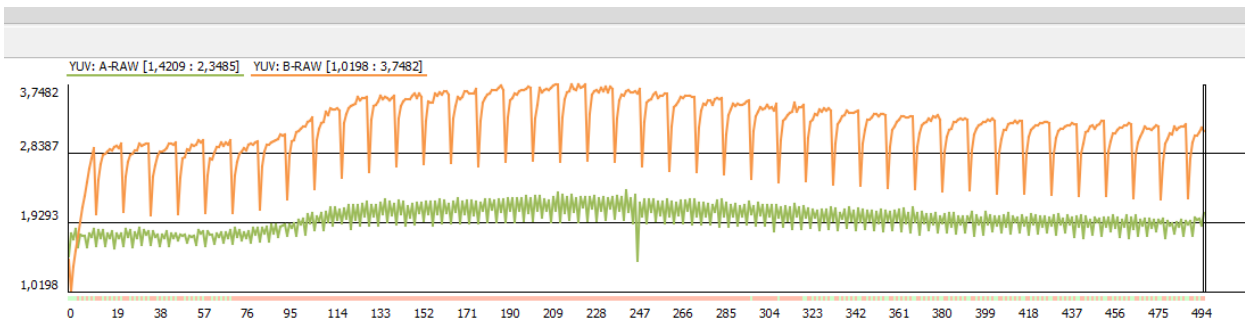


Рисунок 4.9 – Значення метрики VQM для кадрів послідовності Качки

Висновки до розділу

Після аналізу об'єктивних метрик якості відео для кодеків H.264 та MPEG-2 можна зробити висновок, що H.264 виявився більш ефективним у порівнянні з MPEG-2. У випадку тестових зображень "Качки" та "Парк" H.264 продемонстрував вищі значення метрик якості, таких як PSNR та SSIM.

Зокрема, високе значення PSNR (33,605 та 34,217) та SSIM (0,90024 та 0,91875) для H.264 свідчать про високу якість відео та високу структурну схожість з оригіналом. У порівнянні, MPEG-2 має нижчі значення PSNR (29,437 та 29,069) та SSIM (0,77654 та 0,75842), що вказує на меншу якість та меншу структурну схожість.

Низьке значення MSAD (середнє 3,6966) для H.264 свідчить про точне відтворення деталей, що важливо для візуального враження глядача. З іншого боку, високі значення VQM та NQI для MPEG-2 свідчать про високу вартість якості та менш ефективне стиснення даних порівняно з H.264.

Метрика Delta вказує на відмінність між оригінальним відео та його стисненою версією. Низьке значення Delta для H.264 (0,23195 та 0,20792) свідчить про ефективність відтворення даних, тоді як для MPEG-2 воно вище (0,1065 та 0,15043), що може свідчити про більші втрати інформації під час стиснення.

APSNR (Absolute PSNR) визначає відмінність величини PSNR між оригінальним та стисненим відео. Вищі значення APSNR для H.264 (33,605 та 34,217) порівняно з MPEG-2 (29,437 та 29,069) підтверджують його перевагу у збереженні високої якості відео.

MSE (Mean Squared Error) вимірює середньоквадратичну помилку між оригінальним та стисненим відео. Низьке значення MSE для H.264 і високе для MPEG-2 також вказують на більшу точність відтворення зображення H.264.

Висновки з дослідження вказують на вищу ефективність кодеку H.264, який проявив високу якість відео та ефективне стиснення даних. У той час як MPEG-2 виявився менш ефективним у відтворенні високоякісного відео та стисненні даних. Використані об'єктивні метрики, такі як PSNR, MSAD, SSIM, VQM, NQI, допомогли числово оцінити різні аспекти якості відеозображень. З урахуванням цих результатів, важливо враховувати вибір кодеку в залежності від конкретних потреб та умов використання, але H.264 видається більш привабливим варіантом для багатьох мультимедійних застосувань.

Можна визначити конкретні варіанти використання кодеків H.264 та MPEG-2 у різних контекстах:

1. H.264 для стрімінгу та онлайн-відео:

З урахуванням високих значень об'єктивних метрик якості, H.264 може бути ефективним вибором для платформ стрімінгу та онлайн-відео. Висока якість та ефективне стиснення дозволять забезпечити якісне відео при обмежених мережевих ресурсах.

2. MPEG-2 для телебачення та стандартів:

З урахуванням історичного застосування MPEG-2 у телебаченні та стандартах, його можна використовувати там, де цей кодек вже широко застосовується або де низька обчислювальна складність є ключовою. Наприклад, у випадках, коли необхідно дотримуватися існуючих стандартів телевізійного мовлення.

3. Комбінація для різних вимог:

Залежно від конкретних потреб проекту чи сервісу, можливе використання обох кодеків для різних сценаріїв. Наприклад, H.264 може використовуватися для високоякісного стрімінгу, тоді як MPEG-2 може бути використаний у телевізійних системах або для старіших пристроїв.

4. Локальне зберігання та відтворення:

З урахуванням високої ефективності H.264, його можна використовувати для локального зберігання та відтворення відео, де важлива висока якість при ефективному використанні простору зберігання.

ВИСНОВКИ

Завдання дисертаційної роботи дослідити алгоритми стиснення даних для забезпечення ефективності передавання мультимедійної інформації виконано. Під час виконання досліджень з'ясовано, що під час стиснення зображень за стандартом MPEG-4 застосовують більшу деталізацію поля зображення порівняно зі стандартом MPEG-2.

Однією з проблем є втрати, пов'язані зі стисненням, які призводять до різних артефактів та втрат якості відновленого відео. Особливо важливою є проблема артефактів, пов'язаних із часовим стисненням, де надлишковість між послідовними кадрами може впливати на якість відтворення, особливо в ситуаціях зі стрілянням або інтенсивним рухом на екрані.

Також важливою проблемою є нерівномірний рух та обробка кількох незалежних об'єктів на екрані, що може призводити до артефактів, залежних від руху. Це особливо актуально для відео зі спортивних подій, де об'єкти можуть рухатися не прогнозовано та швидко.

Порівнюючи два кодеки, слід відзначити, що H.264 став більш популярним завдяки своїй здатності забезпечити високу якість відео при менших розмірах файлів, його широко застосовують у сучасних технологіях, таких як стрімінгове відео, відеоконференції тощо.

Іншим важливим аспектом є застосування різноманітних технік компенсації руху та інтерполяції для прогнозування кадрів, що сприяє зменшенню обсягу даних при збереженні достатньої якості зображення. Техніки цілочислового та дробового векторного руху відіграють ключову роль у забезпеченні високоефективного кодування руху та зменшенні бітової швидкості.

Об'єктивні методи вимірювання якості можна класифікувати за обчислювальною складністю та ступенем відповідності суб'єктивним оцінкам. Перша група методів, яка включає PSNR, MSAD, Delta, BFM, DFM, NE, SCD, APSNR, MSE, NQI, характеризується низькою обчислювальною складністю і, в основному, і вони не враховують особливостей зору людини. Друга група, до якої

відносяться SSIM, MS-SSIM, VQM, PQR має середню складність та враховує деякі аспекти зору людини. Третя група, яка включає VQM та PQR, є методами високої обчислювальної складності та враховує багато аспектів візуального сприйняття.

Після аналізу об'єктивних метрик якості відео для кодеків H.264 та MPEG-2 можна зробити висновок, що H.264 виявився більш ефективним у порівнянні з MPEG-2. У випадку тестових зображень "Качки" та "Парк" H.264 продемонстрував вищі значення метрик якості, таких як PSNR та SSIM.

Зокрема, високе значення PSNR (33,605 та 34,217) та SSIM (0,90024 та 0,91875) для H.264 свідчать про високу якість відео та високу структурну схожість з оригіналом. У порівнянні, MPEG-2 має нижчі значення PSNR (29,437 та 29,069) та SSIM (0,77654 та 0,75842), що вказує на меншу якість та меншу структурну схожість.

Обидва кодеки і MPEG-2 і MPEG-4 забезпечують широкі можливості щодо формування інтенсивності сигнального потоку стисненого зображення за рахунок вибору параметрів стиснення. Кодек MPEG-4 забезпечує більш ефективне стиснення, але потребує більше обчислювальних ресурсів. Кодек MPEG-2 доцільно використовувати з максимальними параметрами якості для систем цифрового кінематографу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Попович, П. В. Удосконалення методів керування відеопотоком у системах цифрового телебачення за критерієм якості відеозображення URL: <https://ela.kpi.ua/handle/123456789/19323> (дата звернення 26.09.2023).
2. Video Quality measurement tool designed to compare quality of encoded streams based on objective metrics URL: <https://www.elecard.com/products/video-analysis/video-quality-estimator> (дата звернення 26.09.2023).
3. A Guide to MPEG Fundamentals and Protocol Analysis URL: http://www.img.lx.it.pt/~fp/cav/Additional_material/MPEG2_overview.pdf (дата звернення 26.09.2023).
4. Xiph.org Video Test Media URL: <https://media.xiph.org/video/derf/> (дата звернення 26.09.2023).
5. Ultra Video Group Dataset URL: <http://ultravideo.fi/#testsequences> (дата звернення 26.09.2023).
6. Janusz Klink, Tadeus Uhl “Video Quality Assessment: Some Remarks on Selected Objective Metrics” URL: <https://ieeexplore.ieee.org/document/9238303> (дата звернення 26.09.2023).
7. Image Quality Assessment: From Error Visibility to Structural Similarity URL: https://www.researchgate.net/publication/3327793_Image_Quality_Assessment_From_Error_Visibility_to_Structural_Similarity (дата звернення 26.09.2023).
8. Toward A Practical Perceptual Video Quality Metric, Netflix Technology Blog URL: <https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652> (дата звернення 26.09.2023).
9. Критичні значення коефіцієнта лінійної кореляції Пірсона URL: http://psystat.at.ua/Articles/Table_Pearson.PDF (дата звернення 26.09.2023).
10. ITU-R Recommendation BT.500-13. Methodology for the subjective assessment of the quality of television pictures URL: https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-RECBT.500-12-200909-S!!PDF-E.pdf (дата звернення 26.09.2023).
11. Elecard Video Quality Estimator 2023 User Guide URL: https://www.elecard.com/storage/docs/EVideoQuEst_UG.pdf (дата звернення 26.09.2023).
12. VQEG Objective Video Quality Model Test Plan URL: Режим доступу до інформації: ftp://ftp.crc.ca/crc/vqeg/phase1-docs/phase1_obj_test_plan.pdf (дата звернення 26.09.2023).

ДОДАТОК А

Результати вимірювання об'єктивних метрик PSNR, VQM та MS-SSIM для тестових відеопослідовностей у програмі Elecard Video Quality Estimator

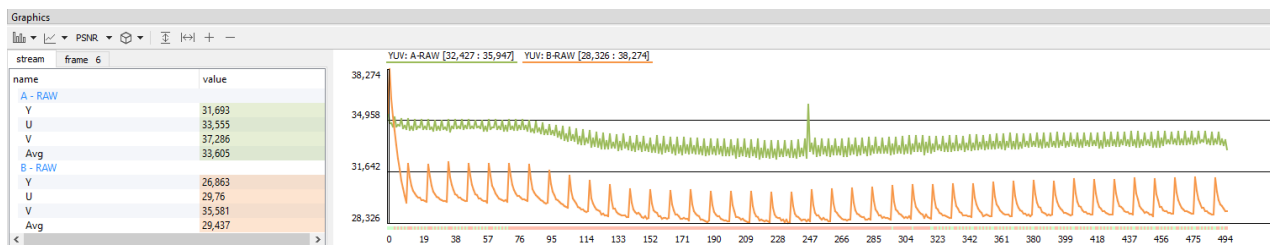


Рисунок А.1 – Значення метрики PSNR для кадрів послідовності Качки

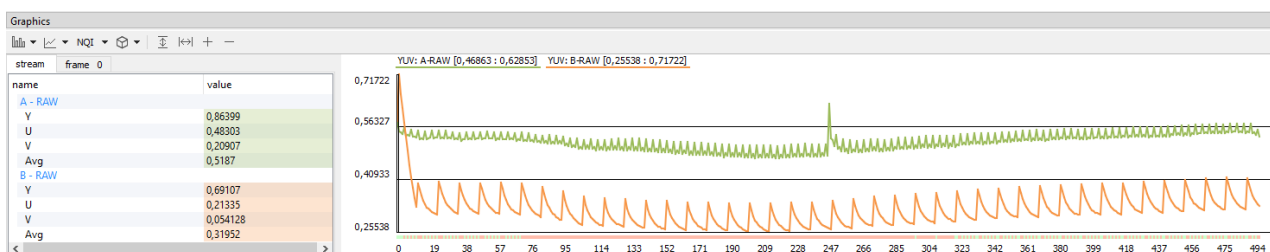


Рисунок А2. – Значення метрики NQI для кадрів послідовності Качки

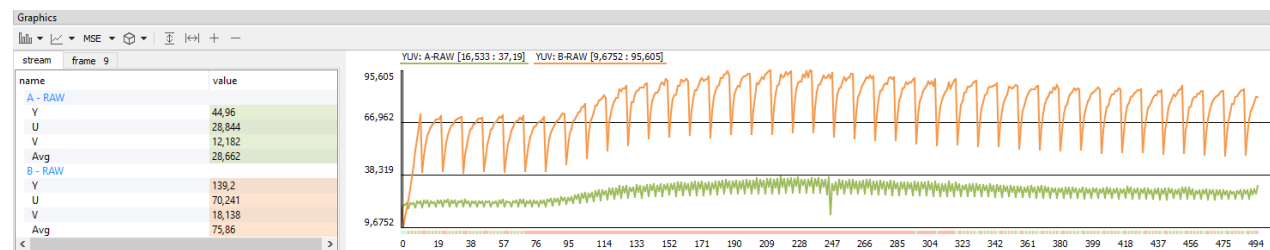


Рисунок А.3 – Значення метрики MSE для кадрів послідовності Качки

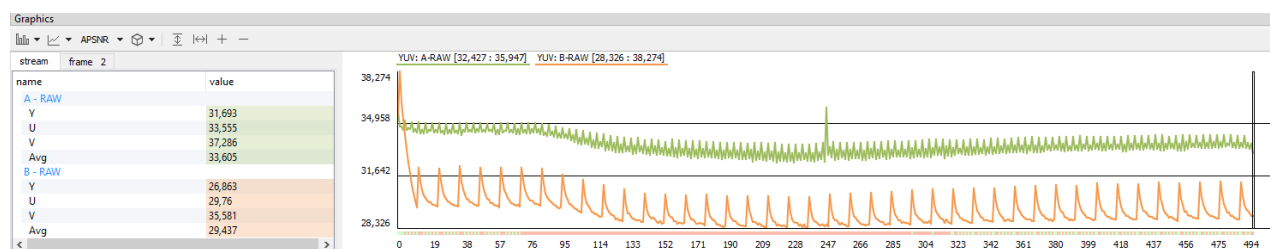


Рисунок А.4 – Значення метрики APSNR для кадрів послідовності Качки

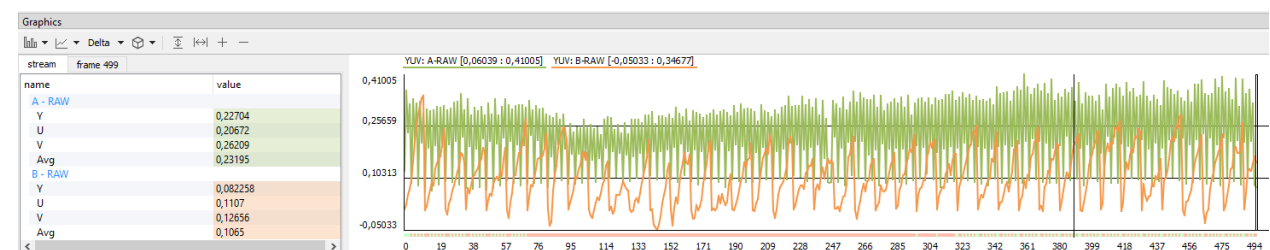


Рисунок А.5 – Значення метрики Delta для кадрів послідовності Качки

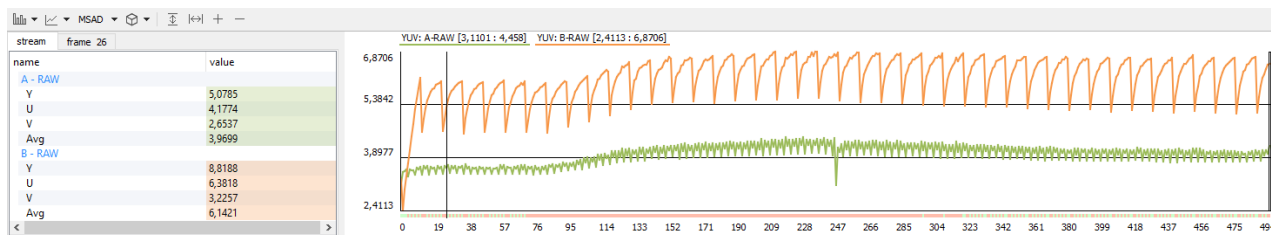


Рисунок А.6 – Значення метрики MSAD для кадрів послідовності Качки

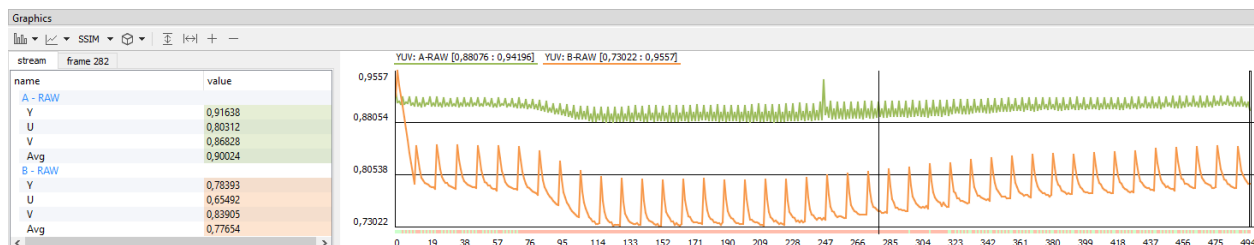


Рисунок А.7 – Значення метрики SSIM для кадрів послідовності Качки

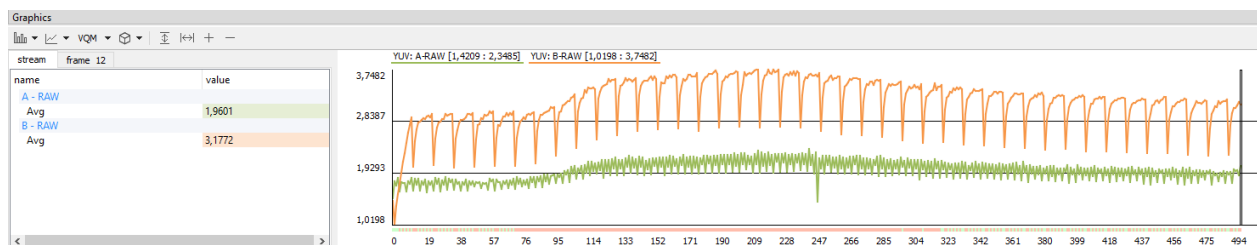


Рисунок А.8 – Значення метрики VQM для кадрів послідовності Качки

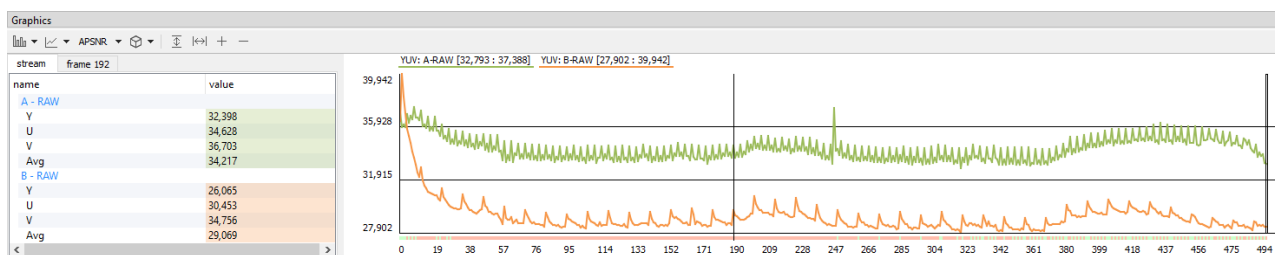


Рисунок А.9 – Значення метрики APSNR для кадрів послідовності Парк

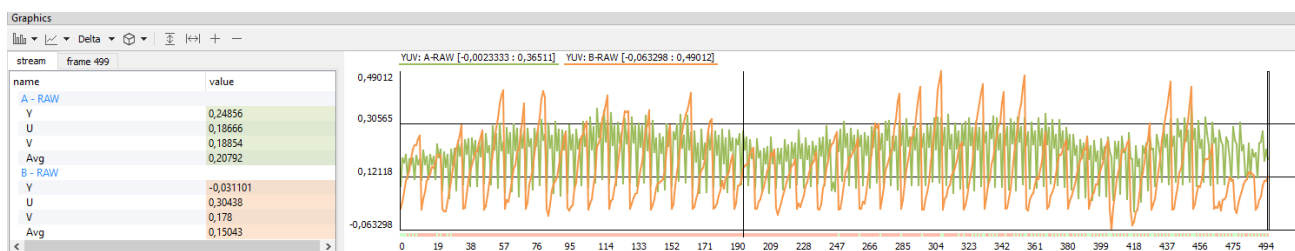


Рисунок А.10 – Значення метрики Delta для кадрів послідовності Парк

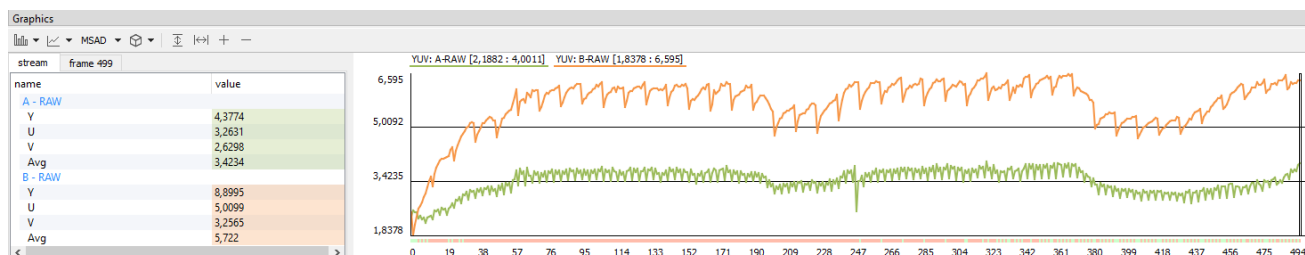


Рисунок А.11 – Значення метрики MSAD для кадрів послідовності Парк

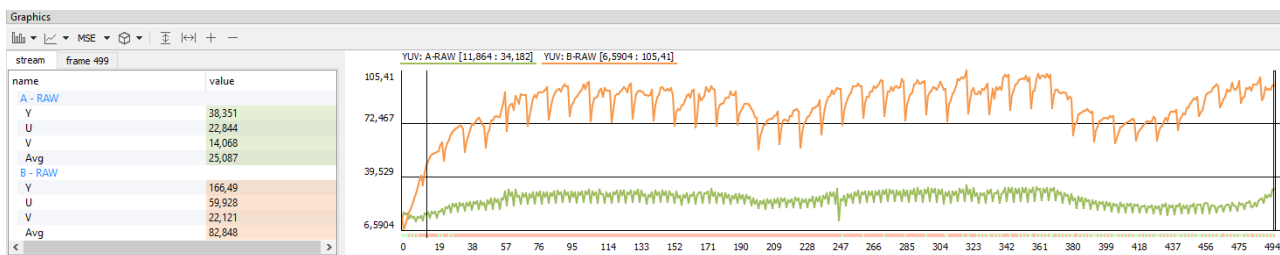


Рисунок А.12 – Значення метрики MSE для кадрів послідовності Парк

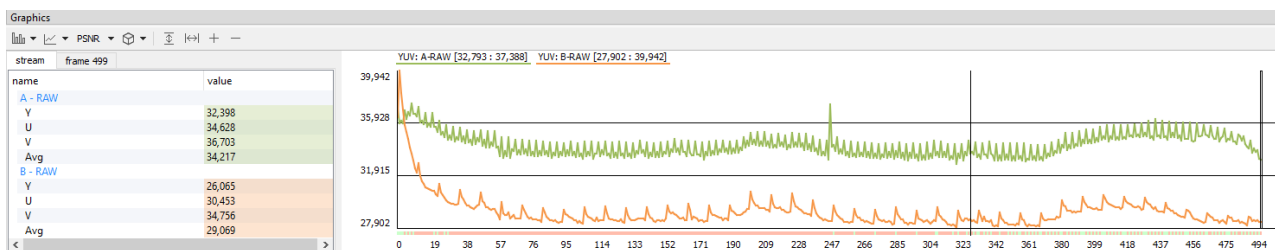


Рисунок А.13 – Значення метрики PSNR для кадрів послідовності Парк

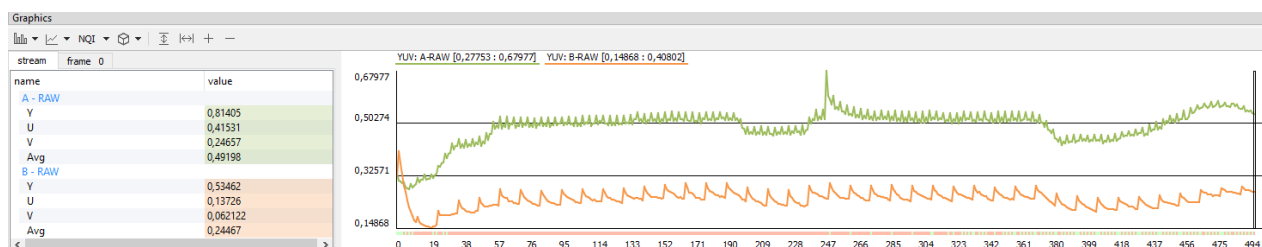


Рисунок А.14 – Значення метрики NQI для кадрів послідовності Парк

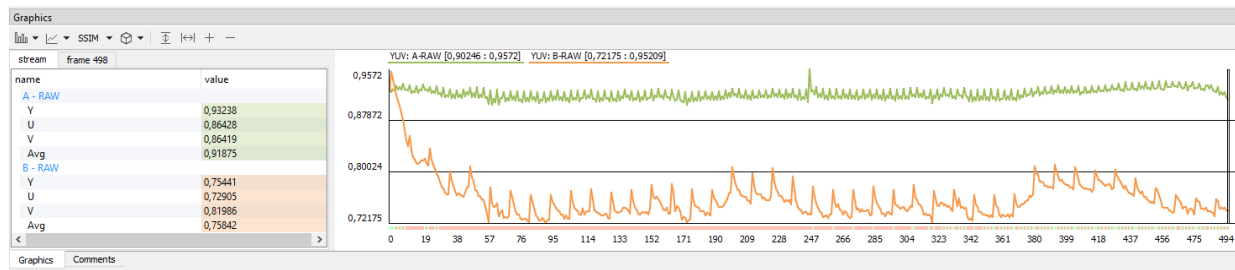


Рисунок А.15 – Значення метрики SSIM для кадрів послідовності Парк

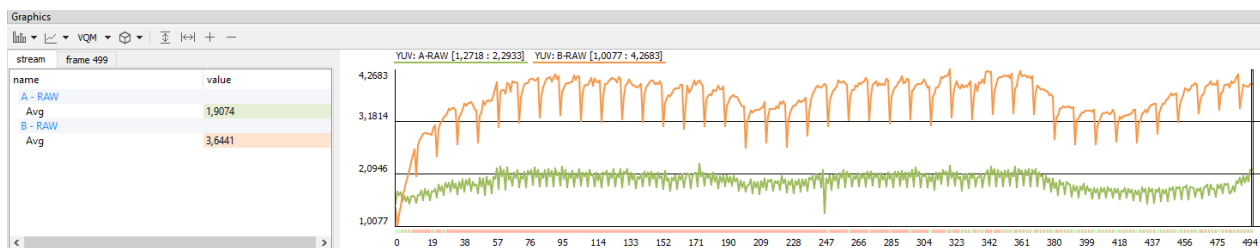


Рисунок А.16 – Значення метрики VQM для кадрів послідовності Парк