

КАПИТЕЛ 5 / CHAPTER 5 <sup>27</sup>EFFICIENT METHODS OF INTELLECTUAL IMAGE AND VIDEO  
PROCESSING BASED ON VISUAL QUALITY METRICS FOR EMERGING  
APPLICATIONS

DOI: 10.30890/2709-2313.2023-21-01-012

**Вступ**

В рамках виконання проекту в 2021 році головну увагу було приділено розробці комбінованих метрик для нових застосувань та методів обробки зображень з урахуванням метрик візуальної якості. Зокрема, розглянуто питання розробки та оптимізації комбінованої метрики з еталоном для зображень, що спотворені множинними викривленнями (одночасно двома чи більшою кількістю типів викривлень) [1]. Також проаналізовано можливості контролю якості триканальних зображень дистанційного зондування, що стиснуті відомим кодером BPG (better portable graphics) [2]. Розроблено комбіновану метрику якості панорамних зображень, яка використовує ентропійні ознаки [3]. Досліджено можливості швидкого визначення ступеня складності зображення та його стиснення з втратами із більш точним забезпеченням бажаної якості [4]. Почнемо з більш простих застосувань, як-то стиснення зображень з втратами, і продовжимо більш складними питаннями оптимізації комбінованих метрик для більш специфічних застосувань.

**5.1. Стиснення зображень кодером BPG із забезпеченням бажаної якості**

Загальна тенденція сучасного дистанційного зондування (ДЗ) — отримання даних з більш високою роздільною здатністю, для більших територій і частіше [5]. Це призводить до збору даних дуже великого розміру і до типових проблем великих даних, коли виникають труднощі з обробкою, передачею, зберіганням і поширенням зображень [6]. Труднощі з передачею, зберіганням і поширенням даних ДЗ по лінії зв'язку можна частково вирішити шляхом їх стиснення [7]. Як відомо [8], існують методи стиснення зображень без втрат і з втратами. Перша група методів дозволяє отримати вихідні неспотворені дані після декомпресії.

<sup>27</sup>Authors: Lukin V.V., Li F., Abramov S.K., Rubel A.S., Okarma K., Lech P., Chlewicki W., Kopytek M.



Проте коефіцієнт стиснення (КС, compression ratio - CR) є невеликим (обмеженим) і досягає значень близько 5 лише у випадках стиснення гіперспектральних зображень, для яких спостерігається дуже висока міжзонна (міжканальна) кореляція [9]. На практиці може знадобитися більший КС, і тоді єдиним виходом є застосування стиснення з втратами, для якого спотворення неминучі [10].

Але рівень спотворень теж не може бути занадто високим. Тоді виникає завдання – забезпечити потрібну якість досить швидко і точно (якість стиснутих зображень може сильно відрізнятись для даного кроку квантування або bpp використовуваного кодера). Одним із виходів є застосування ітераційної процедури, яка виконує багаторазове стиснення/декомпресію, оцінку якості зображення та зміну параметрів кодера [11, 12]. Однак така процедура може вимагати значної та апріорі непередбачуваної кількості ітерацій і, отже, займати багато часу. Альтернативою є застосування двоетапних процедур [13-17], які використовують криві середньої швидкості-спотворення, отримані заздалегідь, щоб визначити параметр кодера на першому кроці, одне стиснення/декомпресію для визначення якості конкретного зображення, яке потрібно стиснути, корекцію параметрів кодера і другий (останній) етап стиснення. Недоліком цієї процедури є те, що іноді її точність у забезпеченні якості стисненого зображення може бути недостатньою [15, 17]. Це трапляється, якщо характеристики конкретного зображення сильно відрізняються від «середніх». Тоді можна провести просту класифікацію зображень перед їх стисненням із втратами [14] і використовувати різні середні криві для різних груп зображень. Це частково допомагає, але можуть бути «дивні» зображення. Крім того, бажано провести попередню класифікацію швидше і надійніше.

Нижче ми показуємо, що класифікацію складності зображень можна зробити на основі обчислення ентропії та порівняння з набором порогових значень. Це дозволяє швидко виявляти «дивні» зображення та розділяти інші зображення на три групи (прості, середньої та високої складності). Це призводить до кращого (точнішого) забезпечення бажаного пікового відношення сигнал-шум (ПВСШ, peak signal-to-noise ratio – PSNR) в діапазоні, де внесені спотворення є візуально помітними, тобто для головного робочого діапазону.

Для керування КС та якістю використовуються різні параметри стиснення з втратами залежно від того, чи є цей кодер на основі дискретного косинусного перетворення (ДКП, discrete cosine transform – DCT), дискретного ієйвлетного



перетворення (ДВП, discrete wavelet transform – DWT) чи іншого перетворення. У кодері AGU на основі ДКП [18], що розглядається в цьому розділі, параметром є крок квантування (КК, quantization step - QS). У попередніх роботах [13, 15] нами вивчено залежність візуальної якості від QS, а також досліджено декілька візуальних показників. Було продемонстровано, що існують загальні тенденції, коли більший QS відповідає нижчій візуальній якості для кожного окремого зображення. Однак значення метрики для різних зображень сильно відрізняються для даного QS, особливо для PSNR і особливо якщо PSNR становить близько 30 дБ. Тому точність двоетапного методу для PSNR нижча, ніж для інших метрик, і може бути навіть незадовільною.

Для підвищення точності двоетапного методу стиснення із втратами в [14] було введено поняття складності зображень. Зображення з інтенсивною текстурою часто характеризується нижчим PSNR, ніж просте зображення для того ж самого QS, і це означає, що складність зображення є ще одним фактором, який визначає якість зображення. Отже, запропонований підхід полягав у тому, щоб розділити зображення на прості та складні структурні та мати відповідно дві осереднені криві КС-спотворення. Стратегія класифікації в [14] полягала в тому, щоб спрогнозувати значення PSNR для набору зображень для деяких значень QS або обчислити КС для стиснення без втрат ( $CR_{\text{lossless}}$ ). Передбачалося, що ці параметри, порівняні з порогами, здатні класифікувати зображення відповідно до їх складності. Розрахунок параметру  $CR_{\text{lossless}}$  займає багато часу в порівнянні з двоетапним методом. Тому значення PSNR для QS, що дорівнює 30, можна використовувати для заміни  $CR_{\text{lossless}}$ , але це також вимагає додаткових обчислень. Адаптивний двоетапний метод реалізовано таким чином: спочатку проводився алгоритм прогнозування PSNR на тестовому зображенні, потім класифікація; нарешті, для виконання двоетапної процедури вибирається відповідна осереднена крива КС-спотворення. Цей підхід частково допомагає підвищити точність, але прогноз PSNR все ще не є швидким, як було б бажано. Інша проблема полягає в тому, що «екстремальні» зображення важко вказати (виявити), що є недоліком для двоетапного методу.

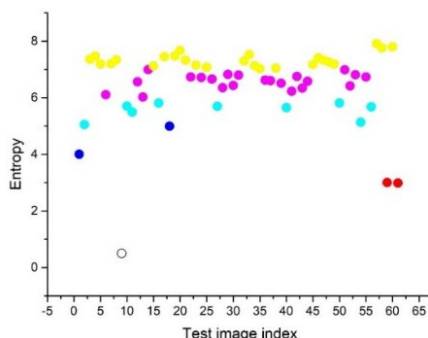
У цьому розділі ми пропонуємо новий підхід до вирішення зазначених вище проблем. Наша ідея заснована на припущенні, що ентропія здатна характеризувати складність зображення. Фактично, ентропія була використана раніше для визначення складності зображень для виявлення дефектів на 3D-друкованих поверхнях [19]. Для перевірки теоретичної доцільності були



отримані дані ентропії для нашого набору зображень (включаючи 61 зображення у відтинках сірого розміром 512×512 пікселів). Результат показаний на рисунку 1.

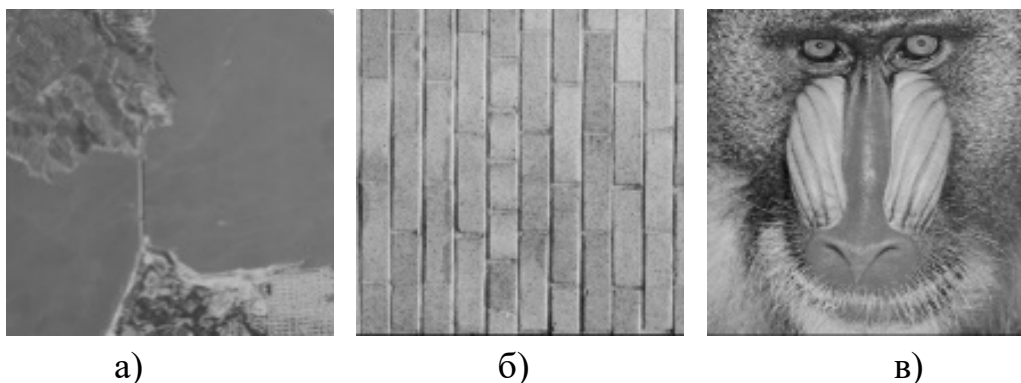
Для 61 дослідженого зображення ентропія змінюється в межах від 0,5 до 7,9, а  $CR_{\text{lossless}}$  – від 100,7 до 1. Коефіцієнти кореляції для цих двох показників складності зображення становлять -0,74 для кореляції Пірсона, -0,88 для рангової кореляції Спірмена та -0,76 для рангової кореляції Кендалла. Більший  $CR_{\text{lossless}}$  відповідає простішим зображенням; при цьому значення ентропії менше, і навпаки. Залежно від значення ентропії зображення легко поділити набір зображень на три групи:

- а) складне зображення, якщо ентропія  $>7$ ;
- б) зображення середньої складності, якщо  $6 < \text{ентропія} \leq 7$ ;
- в) просте зображення складності, якщо ентропія  $< 6$ .



**Рисунок 1 – Значення ентропії для 61 тестових зображень в градаціях сірого**

На рисунку 2 показано три приклади, що відповідають трьом рівням складності.



*а) низька складність; б) середня складність; в) висока складність.*

**Рисунок 2 – Ілюстрація рівнів складності зображень**



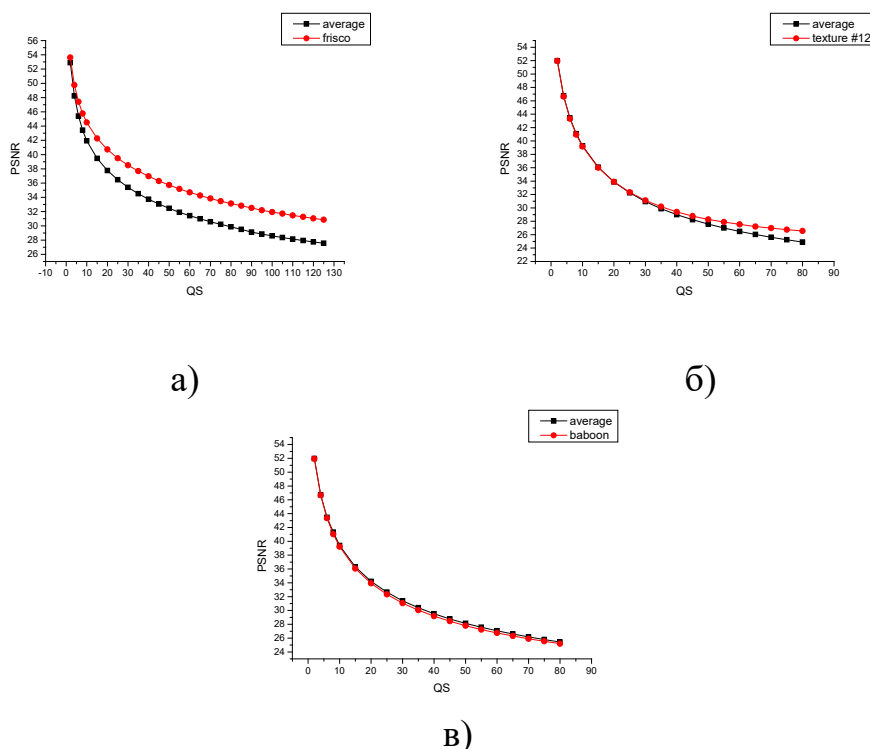
На рисунку 2,а) показано зображення ДЗ Фріско, яке має велику однорідну ділянку (фон), і воно належить до простого набору зображень відповідно до ентропії (5,8166). На рисунку 2,б) показано зображення текстури (пронумеровано як #12 у нашому наборі), яке має чітку текстуру, і воно належить до набору зображень середньої складності, оскільки його ентропія дорівнює 6,799. Останнє — це зображення Baboon з ентропією, що дорівнює 7,3579, що є типовим складним зображенням через його багату інформацію про текстуру та складну композицію. На основі порівняння трьох зображень це доводить, що наша стратегія класифікації приблизно відповідає складності зображення.

Аналогічно [14], осереднені криві КС/спотворення були отримані для трьох рівнів базових наборів зображень, 4 зображення у простому наборі зображень, 15 зображень у наборі зображень середньої складності та 17 зображень у наборі складних зображень. Ці криві показані на рисунку 3; криві для трьох зображень, що показані на рисунку 2, також наведені на відповідних графіках. Три осереднені криві мають спільну тенденцію, PSNR монотонно зменшується разом із збільшенням QS, при цьому ці осереднені криві приблизно «локально паралельні» кривим для зображень відповідної складності, але відхилення кривизни різних наборів зображень є різними. Серед них зображення в простому наборі зображень мають найбільшу різноманітність порівняно з двома іншими наборами.

Перший крок стиснення виконується за допомогою  $QS_{init}$ , потім розпаковується та обчислюється візуальна якість розпакованого зображення щодо вихідного зображення. Цей початковий PSNR ( $PSNR_{init}$ ) зазвичай досить близький до бажаного  $PSNR_{des}$ , але його може бути недостатньо, щоб задовольнити вимоги користувача; отже, другий крок може бути необхідним для зменшення похибки. QS коригується рівнянням (2) за допомогою  $PSNR_{init}$ . Другий крок виконується з виправленим  $QS(QS_{des})$  і створюється стиснутий файл зображення, який забезпечує PSNR ( $PSNR_{pro}$ ) ближче до  $PSNR_{des}$ , ніж  $PSNR_{init}$

$$QS_{des} = QS_{init} + \frac{PSNR_{des} - PSNR_{init}}{M'} \quad (2)$$

Тестові експерименти були проведені на 61 зображенні (включаючи 36 основних зображень), а статистичні результати наведені в таблиці 1, де як бажана якість були вибрані три значення PSNR (40, 35 і 30 дБ). Ми визначили дисперсію  $PSNR_{init}$  для першого кроку (позначена як  $VAR_{fir}$ ) і дисперсію  $PSNR_{pro}$  для другого кроку (позначена як  $VAR_{sec}$ ),  $MAX_{\Delta final}$  позначає максимальну абсолютну



а) для простих зображень; б) для зображень середньої складності;  
в) для складних зображень.

### Рисунок 3 – Осереднені та конкретні залежності ППСШ від КК

помилку між  $PSNR_{pro}$  та  $PSNR_{des}$  у групі даних. Результат доводить, що стиснення на другому етапі ефективно зменшило похибку; дисперсія  $PSNR_{pro}$  не перевищує  $1,5 \text{ дБ}^2$ . Порівняно з основним двоетапним методом, точність, загалом, була підвищена завдяки нашому новому підходу на основі ентропії. Проте є деяка детальна інформація, на яку потрібно звернути увагу. Похибки в простому наборі зображень більші, ніж в інших двох наборах, і найбільша похибка в ньому виникає в бажаному PSNR, що дорівнює 40 дБ, тоді як в інших двох наборах найбільша похибка виникає в найнижчому бажаному PSNR (як це було в наших роботах [13-15]).

Проблеми для простого набору зображень пов'язані з трьома «дивними» зображеннями, представленими на рисунку 4: наведено залежності PSNR від QS. Як бачимо, це штучно створені зображення з обмеженою кількістю рівнів значень (що показали гістограми значень зображень). Криві КС-спотворення для них можуть бути немонотонними або сильно відрізняться від осередненої кривої. Це викликає проблеми з використанням виразів (1) і (2) для забезпечення бажаної якості. Тоді попередній розрахунок ентропії може показати на практиці, що маєш справу з «дивним» зображенням (якщо ентропія менша за 3, «будьте





обережні»).

**Таблиця 1 – Результати експериментів**

Набори зображень	$PSNR_{des}$	$VAR_{fis}$	$VAR_{sec}$	$MAX_{\Delta final}$
Прості	40	26,1129	1,4352	4,4539
	35	20,153	1,4246	2,8279
	30	15,2624	1,1545	1,9385
Середньої складності	40	0,1583	0,0033	0,2394
	35	0,6986	0,486	0,8277
	30	2,4614	0,4111	1,9070
Складні	40	0,3173	0,0238	0,5398
	35	1,3866	0,1221	1,4162
	30	3,6813	0,3083	2,4022

Таким чином, розглянуто можливість забезпечення бажаної якості стиснутих зображень із втратами (що характеризується PSNR) для кодера AGU на основі ДКП. Продемонстровані основні проблеми, пов'язані з тим, що складність зображення, підданого стисканню, може сильно відрізнятися.



а) фактично чорно-біле зображення; б) штучне зображення;  
в) штучне зображення.

**Рисунок 4 – Приклади «дивних» зображень**

У зв'язку з цим характеристики КС-спотворення для різних зображень сильно відрізняються. Дотримуючись раніше запровадженого підходу, нами запропоновано провести попередню класифікацію складності зображень, що дозволяє розрізняти зображення на прості, середні та складні (а також виявляти «дивні» зображення). Перевага запропонованого підходу до класифікації полягає в тому, що він базується на ентропії, яку можна швидко обчислити, а потім порівняти з пороговими значеннями. Оскільки попередньо визначено декілька осереднених кривих КС-спотворення, можливе застосування двоетапного методу стиснення з використанням результатів попередньої класифікації. Це



дозволяє значно підвищити точність у забезпеченні бажаної якості стиснутих зображень.

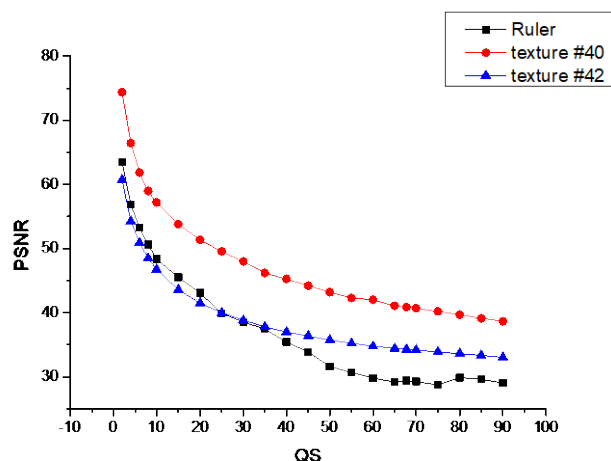


Рисунок 5 – Залежності ПКСШ від КК для «дивних» зображень

## 5.2. Забезпечення бажаної якості для стиснення з втратами кодером ВРГ для триканальних зображень дистанційного зондування

Існує багато різних вимог до стиснення з втратами. Їх пріоритет і тип (обов'язкова або бажана вимога) залежать від застосування [20]. Може бути пріоритетна вимога надання КС не менше необхідного значення. Це відбувається, якщо пропускна здатність каналу та час передачі даних обмежені (фіксовані). У такій ситуації кодер повинен мати простий алгоритм забезпечення бажаного КС (як методи, подібні до JPEG2000) і зменшувати внесені спотворення, щоб запобігти їх негативному впливу на подальшу обробку зображення. Якщо мати справу з багатоканальними даними (наприклад, кольоровими, мультиспектральними, гіперспектральними), то в такому випадку попередня декореляція та використання тривимірних (3D) підходів до стиснення зазвичай покращують ефективність стиснення [21]. Однак може бути й інший пріоритет вимог. Можна побажати, щоб спотворення (що характеризуються певною мірою чи метрикою) були менше (або не гірше), ніж заданий рівень, з одночасним бажанням мати якомога більший КС. Тут розробник стикається з кількома проблемами. Однією з них є проблема швидкого та ефективного контролю (забезпечення) рівня спотворення [22]. Інша проблема полягає в тому, яку метрику вибрати, щоб характеризувати спотворення адекватно та з урахуванням завдання, яке слід розв'язувати за допомогою стиснених даних [23].





Можуть бути встановлені додаткові вимоги. Наприклад, стиснені дані мають бути представлені в певному форматі (за деякими стандартами) або стиснення повинно виконуватися дуже швидко та з обмеженим споживанням енергії [24].

Щоб задовольнити ці вимоги, останнім часом з'явилися (чи були створені) деякі передумови. По-перше, значна увага була приділена метрикам, здатним характеризувати якість даних ДЗ [23, 25, 26]. Нами було продемонстровано [23], що індекс подібності середнього відхилення (Mean Deviation Similarity Index - MDPI) [27] та деякі інші елементарні метрики візуальної якості можуть добре характеризувати якість трьохканальних зображень ДЗ зі спотвореннями, типовими для зображень дистанційного зондування, включаючи спотворення внаслідок стиснення з втратами. По-друге, було показано, що стиснення з втратами за певних умов може призвести до практично такої ж або навіть кращої ефективності класифікації зображень порівняно з класифікацією вихідних (нестиснутих) даних [28-30]. Це відбувається, якщо спостерігається деяке придушення шуму або якщо спотворення неможливо виявити візуально [31]. Це означає, що можна забезпечити одночасно дві переваги –отримати КС, який достатньо відрізняється від одиниці (суттєво перевищує 1), і спостерігається покращена (або, принаймні, не гірша) класифікація. По-третє, була проведена достатня робота для прискорення стиснення з втратами з досягненням попередньо визначеної (бажаної) якості. Ітераційне стиснення [32] здатне забезпечити високу точність, але вимагає непередбачуваної кількості ітерацій, що може спричинити проблеми з часом та обчислювальною ефективністю стиснення. Натомість були запропоновані та досліджені двокрокові методи та алгоритми [13-17]. Було продемонстровано, що стиснення з втратами із забезпеченням заданої якості відповідно до вибраної метрики якості може бути здійснено з точністю, задовільною для практики, у два етапи. Для першого кроку потрібна осереднена залежність вибраної метрики від параметра, який керує стисненням (parameter that controls compression - PCC), що отримана заздалегідь для набору базових зображень (див. попередній розділ та нижче у даному розділі). Маючи в розпорядженні таку залежність, легко визначити початковий PCC, який відповідає бажаному значенню метрики якості відповідно до кривої середньої швидкості спотворення (наприклад, пікове співвідношення сигнал/шум (PSNR) від кроку квантування (QS)) або біт на піксель (bits per pixel - bpp)). Стискаючи та розпаковуючи зображення за допомогою цього початкового PCC, можна отримати значення метрики для першого кроку. Після



цього РСС коригується за допомогою лінійної інтерполяції та осередненої кривої КС-спотворення, щоб отримати остаточний РСС, який буде використаний для остаточного стиснення на другому кроці.

Така двоетапна процедура є досить швидкою, точною й універсальною. Вона добре працює для кодерів на основі дискретного косинусного перетворення (DCT) і вейвлетів (наприклад, SPIHT). Тут важливо, що двоетапний підхід добре працює для кодеру VPG – нового методу стиснення, який має тенденцію замінити JPEG через значно кращу ефективність порівняно з JPEG, JPEG2000 та деякими іншими популярними методами стиснення з втратами [33]. Характеристики стиснення можна змінювати за допомогою так званого параметра якості Q, його збільшення призводить до більшого CR, але більше спотворень, що вносяться для VPG [33]. Однак одне й те саме Q призводить до різної якості, якщо вона характеризується певною метрикою якості [32], наприклад, PSNR або метрикою PSNR-HVS-M [34], де HVS означає систему зору людини (human vision system), а M означає маскування. Це означає, що Q слід регулювати залежно від зображення, яке підлягає стисненню. У статті [32] показано, що це можливо, якщо потрібно стиснути зображення у відтінках сірого. Однак у практиці дистанційного зондування багато сучасних систем отримують (формують) багатоканальні зображення, для яких доцільно використовувати тривимірне стиснення. Таким чином, перша мета нашого дослідження полягає в тому, щоб перевірити, чи можна застосувати двоетапний метод для стиснення багатоканальних зображень ДЗ, точніше, триканальних зображень, які включають кольорові зображення та дані про видимий діапазон мультиспектральних зображень. Друга мета полягає в тому, щоб дослідити деякі важливі властивості метрики MDSI і перевірити, чи можуть її бажані значення бути забезпечені двоетапним підходом з відповідною точністю. Зауважимо, що ми обмежені тим фактом, що не існує загальноприйнятих показників якості для довільної кількості компонентів даних ДЗ.

Розглянемо властивості триканальних зображень ДЗ. Як зазначалося вище, багато сучасних систем ДЗ формують багатоканальні зображення, де термін «багатоканальний» відноситься до мультиспектральних, гіперспектральних, двота мультиполяризаційних радіолокаційних даних тощо. Зосередимося на трьохканальних зображеннях через три основні причини. По-перше, разом з радіолокаційними даними з подвійною поляризацією, триканальні зображення є найпростішим прикладом багатоканальних даних ДЗ. Маючи деякі методи та



результати, отримані для трьохканальних зображень, їх можна узагальнити для зображень з більшою кількістю каналів. По-друге, триканальні зображення ДЗ можна візуалізувати як кольорові, що спрощує візуальний аналіз порівняно з випадком представлення багатоканального зображення у псевдокольорах. Хоча довжини хвиль каналів мультиспектральних зображень зазвичай не збігаються з довжинами хвиль у традиційних представленнях RGB, візуалізовані триканальні зображення зазвичай виглядають більш-менш зрозумілими для спостерігача. По-третє, ми вважаємо, що VRG потенційно можна модифікувати для стиснення 3D-зображень з більшою кількістю каналів, ніж три. Однак на даний момент у нас є тільки версія VRG, призначена для стиснення кольорових зображень.

У порівнянні зі звичайними кольоровими зображеннями (фотографіями), зображення ДЗ мають деякі особливості. Перш за все, зображення ДЗ зазвичай більш високоструктуровані і кожен об'єкт має семантичне значення [36], тоді як звичайні зображення більш хаотичні. Ці об'єкти в зображеннях ДЗ потребують аналізу на подальших етапах обробки даних, зокрема, розпізнавання цілей, класифікації, сегментації, оцінки параметрів. Наприклад, головною метою процесу сегментації зображення є розділення зображення на набір сегментів, які є однорідними з точки зору хроматографії або текстури [37]. Між тим, зображення ДЗ часто включають великі ділянки фону, що менш важливо в порівнянні з об'єктами переднього плану [38]. Це можна врахувати при компресії.

По-друге, кореляція між компонентними зображеннями триканальних даних ДЗ може відрізнитися від кореляції компонентів R, G і B кольорових зображень. У той час як для кольорових зображень коефіцієнт взаємної кореляції зазвичай становить близько 0,7 [39], коефіцієнт кореляції між компонентами триканальних даних ДЗ може бути достатньо вищим [9]. Це може вплинути на ефективність стиснення, якщо використовуються тривимірні підходи.

По-третє, зображення компонентів у RGB є 8-бітовими даними. Компонентні зображення ДЗ можуть спочатку бути представлені іншим способом, наприклад, як 16-бітові дані. Тоді важко адекватно порівняти стиснення кольорових зображень та триканальних зображень ДЗ. Через це ми далі будемо розглядати триканальні зображення з 8-бітовим представленням компонентних зображень. Якщо зображення компонентів спочатку мають більше 8 біт, то припустимо, що вони попередньо нормалізовані до 8 біт перед стисненням із втратами. Звичайно, це призводить до внесення додаткових



помилку за рахунок округлення до найближчого цілого числа. Проте внесені помилки відносно невеликі. PSNR зображень внаслідок нормалізації, округлення та повторної нормалізації становить близько 59 дБ. Спотворення через стиснення з втратами зазвичай призводять до меншого PSNR. Таким чином, викривлення внаслідок нормалізації (якщо застосовується) можна ігнорувати.

По-четверте, не існує загальноприйнятих баз даних еталонних (опорних, без спотворень) зображень ДЗ. Більше того, типи спотворення, властиві зображенням ДЗ і кольоровим зображенням, частково збігаються, але й частково відрізняються. Наприклад, згладжування зображення не зустрічається на практиці зображень ДЗ. Між тим, для кольорових зображень спекл не є характерним, але він може бути важливим фактором для зображень ДЗ спеціального типу (радарних із синтезованою апертурою). Це перешкоджає прямому використанню баз даних кольорових зображень для створення висновків і рекомендацій щодо зображень ДЗ. Однак нещодавно база даних TID2013 була опосередковано використана нами для аналізу підмножин типів спотворення, які можуть бути присутніми в зображеннях ДЗ. Це дозволило визначити хороші елементарні метрики візуальної якості для адекватної характеристики якості зображення ДЗ та розробити комбіновані показники. Під елементарною візуальною метрикою якості ми маємо на увазі вже відому метрику, яка використовується в комбінованій метриці як один із вхідних даних. Зокрема, метрика MDSI є однією з найкращих елементарних метрик. Нижче ми надаємо докладнішу інформацію про цей показник і пояснюємо, чому він привернув нашу увагу.

По-п'яте, метрики візуальної якості описують якість даних з певної точки зору. Не зовсім зрозуміло, який зв'язок між показниками візуальної якості та, наприклад, точністю класифікації зображень [28]. У той же час попередні результати класифікації стиснутих зображень вже продемонстрували, що візуальна якість має високу кореляцію з точністю класифікації, особливо для класів, представлених невеликими, протяжними та текстурними об'єктами, тобто для класів, які є досить неоднорідними. Тут важливо зазначити, що високочастотна інформація може бути втрачена через стиснення з втратами для великого КС, і це може негативно вплинути на класифікацію.

MDSI є однією з метрик, що розраховується у порівнянні з еталоном. Найпростішою і класичною метрикою з еталоном є середня квадратична похибка (mean square error - MSE), обчислена шляхом усереднення квадратів різниць



інтенсивності спотвореного та опорного (еталонного) зображення по всім пікселям. Показником, строго пов'язаним з MSE, є пікове відношення сигнал/шум (PSNR). Істотними перевагами MSE і PSNR є те, що їх обчислення є простим, а їх фізичний зміст зрозумілим. Слабка сторона цих показників полягає в тому, що вони не дуже добре підходять до сприйнятої візуальної якості [20]. Через це за останні три десятиліття були запропоновані та інтенсивно вивчені численні метрики, засновані на MSE та інших принципах. наприклад, FSIM [41], SSIM [42, 43], MDSI [27] тощо.

Тож нами вирішено використовувати MDSI як показник візуальної якості для трьохканальних зображень ДЗ. Щоб пояснити причини цього, необхідно згадати деякі аспекти та вимоги до оцінки якості зображень (image quality assessment – IQA). По-перше, якість зображення зазвичай оцінюється в досить складних експериментах з великою кількістю учасників і тестових зображень різної складності. Результатом такого тестування є одержання середньої оцінки (mean opinion score - MOS) або диференційної середньої оцінки (differential mean opinion score - DMOS). Метрика вважається хорошою, якщо для різних баз даних вона має високі абсолютні значення коефіцієнта кореляції між даною метрикою та MOS, де можна брати до уваги як звичайні (Пірсона), так і рангові (Спірмена і/або Кендалла) кореляції (в ідеалі бажано, щоб коефіцієнти кореляції Пірсона і Спірмена мали абсолютні значення, близькі до одиниці). Для оцінювання необхідна монотонність залежності метрики від якості зображення, що характеризується коефіцієнтом кореляції рангового порядку Спірмена (Spearman rank order correlation coefficient - SROCC). Водночас бажаною є і лінійність цієї залежності, яка краще характеризується коефіцієнтом кореляції Пірсона (Pearson correlation coefficient - PCC).

У [23] нами проведені експерименти для 50 відомих метрик. Значення SROCC розраховуються як головний кількісний критерій ефективності метрик для всіх типів спотворень і трьох підмножин бази даних Тампере (TID2013). SROCC для MDSI з MOS становить 0,8897 для всіх типів і рівнів спотворень, що вище, ніж для більшості інших метрик, а SROCC для підмножини Noise&Actual становить 0,9374. Він є найвищим серед усіх розглянутих елементарних метрик. Водночас статистика середнього часу обчислень показує, що обчислювальна ефективність MDSI дуже висока.

По-друге, під час обчислення MDSI величина градієнта використовується для вимірювання структурних викривлень, а кольорові характеристики





використовуються для вимірювання спотворення кольору (нагадаємо, що обидва типи спотворень однаково важливі для трьохканальних зображень ДЗ). Згодом ці дві карти подібності об'єднуються, щоб утворити карту подібності градієнта кольоровості. На відміну від SSIM та FSIM, стратегія об'єднання використовується для обчислення остаточного показника якості. Порівняно з попередніми дослідженнями, ця нова карта подібності градієнта краще співвідноситься з особливостями зору людини.

Забезпечення бажаної візуальної якості під час стиснення з втратами є складним завданням. Може бути корисним, якщо значення метрики пов'язане з певним рівнем якості. Наприклад, було б добре знати діапазон значень показників, для яких спотворення практично не видно.

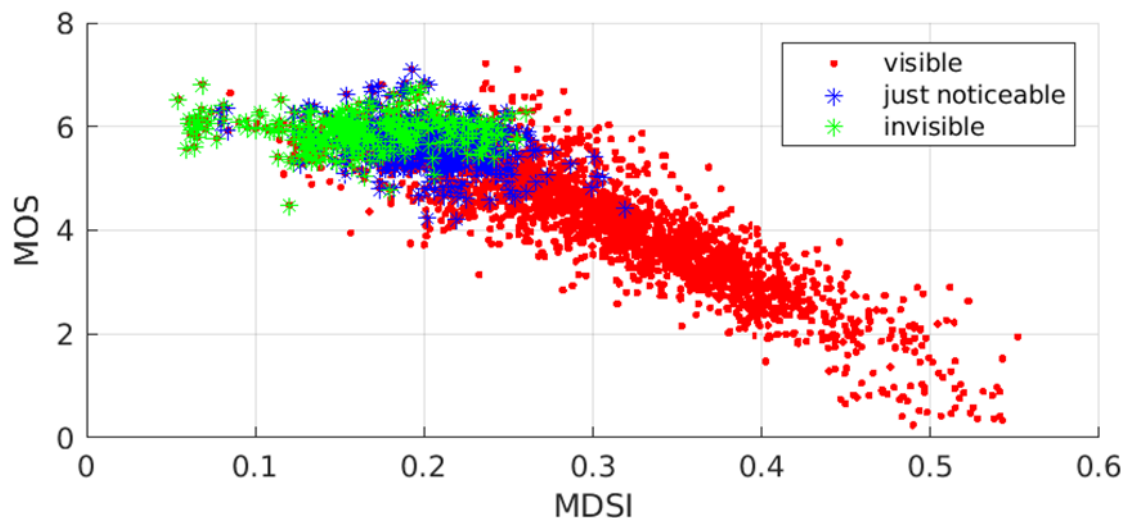
Під час стиснення з втратами бажана візуальна якість часто знаходиться в певному діапазоні. На щастя, раніше нами проведено деякий аналіз на основі інших метрик і вирішено зазначену вище задачу. Нижня межа є такою, що стиснення з втратами не має негативного впливу на подальшу обробку зображення. Це означає, що внесені спотворення можуть бути помітними або навіть видимими, але не драгувати. Що стосується верхньої межі, то стиснення з втратами має забезпечувати вищий КС, ніж стиснення без втрат. Якщо поріг занадто високий, переваги стиснення з втратами зникають. Розумно обраний поріг полягає в тому, що внесені спотворення непомітні. Тоді візуальна якість стиснутих даних буде ідентичною якості стисненого зображення без втрат, але можна досягти більш високого КС. Цей поріг становить близько  $\sim 40$  дБ у відповідності до метрики PSNR-HVS-M для багатоканальних зображень ДЗ. Щоб визначити задовільну область для метрики MDSI, нами протестовано 3000 кольорових зображень бази даних TID2013 для отримання статистичних даних ці зображення були поділені на три категорії: видимі (visible), помітні (just noticeable) та невидимі (invisible) спотворення. MOS в базі було попередньо розраховано для кожного зображення, діаграма розсіювання для MDSI і MOS показана на рисунку 6.

Якість зображень у TID2013 характеризується MOS, що визначається методологією експерименту, проведеного спостерігачами [35]. Поєднуючи статистичні результати та значення MOS, можна приблизно вважати, що існує три градації якості зображення відповідно до MDSI: 1) Відмінна якість ( $MDSI \leq 0,15$ ), спотворення переважно непомітні; 2) Хороша якість ( $0,15 < MDSI \leq 0,25$ ), спотворення можуть бути помітними; 3) середня та погана якість ( $MDSI > 0,25$ ),





спотворення помітні або можуть дратувати. Таким чином, область нашого підвищеного інтересу - від 0,10 до 0,25.



**Рисунок 6 – Діаграма розсіювання (скатерограма) залежності MDSI від MOS**

Двоетапний метод стиснення зображень, як вже зазначалося, спрямований на встановлення того, як контролювати візуальну якість під час стиснення з втратами, а також забезпечити бажану візуальну якість для обробки зображень з точки зору вибраної метрики візуальної якості. Попередні дослідження вже довели, що цей метод добре працює для кодерів на основі ДКП та малохвильових перетворень [13-17].

Згадаємо, що нам потрібно, щоб наш підхід був застосовний до зображень різної місцевості або, іншими словами, різної складності, що містять різні типи об'єктів. Це означає, що в наших експериментах слід використовувати досить багато тестових зображень. Це пояснює методологію проектування та аналізу, використану на різних етапах нашого дослідження.

Перш за все, вибирається певна кількість зображень для стиснення/декомпресії для набору значень РСС; назвемо їх базовим набором зображень. Кожне значення MDSI для кожного тестового зображення, отриманого після стиснення/декомпресії з використанням заданих значень РСС, записується для оцінки спотворень, після чого з цих статистичних даних можна отримати модель залежності метрики візуальної якості від РСС. На цьому етапі можна отримати індивідуальні залежності метрики від РСС. Потім виконується осереднення значення метрики візуальної якості для всіх базових зображень, що



відповідають кожному РСС. Таким чином також можна отримати усереднену залежність, так звану осереднену криву КС-спотворення, яка відображає монотонну зміну якості від РСС. Цей процес виконується в режимі офлайн (до використання для стиснення конкретних зображень), отже, він не впливає на ефективність двоетапного методу у сенсі витрат часу. На основі цієї кривої вже можна застосувати двоетапний метод (приклади розглянуто вище в першому розділі).

Початкове значення РСС визначається з використанням бажаної візуальної якості та осередненої кривої КС/спотворення (або даних у відповідній таблиці). Загалом, воно розраховується наступним чином

$$PCC_{init} = PCC_{est} + \frac{M_{des} + M_{ave}}{M'} \quad (3)$$

де  $M_{des}$  – бажане значення метрики візуальної якості, що встановлюється користувачем, інші три параметри отримують з осередненої кривої, представленої в табличній формі:

$M_{ave}$  – це значення, яке в таблиці є найближчим до  $M_{des}$ ;

$PCC_{est}$  – це значення РСС, що відповідає цьому  $M_{ave}$ ;

$M'$  – похідна кривої для аргументу  $PCC_{est}$ .

Як можна побачити, (3) є узагальненням (1). Розрахунок (3) не включає параметри зображення, яке потрібно стиснути, тому для заданої бажаної візуальної якості значення початкового РСС<sub>init</sub> для будь-якого зображення, яке потрібно стиснути, є однаковим незалежно від властивостей конкретного зображення (в принципі, можливі модифікації, що запропоновані в розділі 1, але вони поки що не вивчалися).

Першим кроком двоетапного методу стиснення (як вже було показано на конкретному прикладі в розділі 1) є стиснення та розпакування зображення з початковим РСС; далі значення метрики візуальної якості розпакованого зображення розраховується у порівнянні з первинним зображенням. Зазвичай значення метрики  $M_{init}$  приблизно дорівнює бажаному  $M_{des}$ , тому надалі потрібно обчислити похибку  $|M_{init} - M_{des}|$  і прийняти рішення, прийнятна ця похибка чи ні, а потім приймати рішення про подальші дії. Для деяких зображень похибка може відповідати необхідній точності, тоді другий крок не потрібен, стиснене зображення на першому кроці стиснення можна розглядати як кінцевий вихід. Якщо ж точність, що характеризується  $|M_{des} - M_{init}|$  не підходить, то потрібен



другий крок стиснення.

Щоб підвищити точність наданої візуальної якості, значення PCC необхідно відкоригувати за допомогою рівняння (4), що є аналогічним (2):

$$PCC_{des} = PCC_{init} + \frac{M_{des} + M_{init}}{M'} \quad (4)$$

Це виправлене значення PCC відрізняється (є індивідуальним) для різних зображень. Стиснення другого етапу здійснюється за допомогою PCC<sub>des</sub>, стиснутий файл зображення другого кроку є кінцевим, тобто стиснення із забезпеченням потрібної якості вважається завершеним.

Оскільки попит на зображення високої якості з прийнятним розміром різко зріс, то в останній час розроблено численні кодери стиснення з втратами [18, 44-46], VP8 – це новий метод стиснення зображень, який має на меті замінити старий формат JPEG. через значно кращу продуктивність у сенсі вищої якості та/або меншого розміру [47]. Незважаючи на те, що нещодавно були запропоновані методи стиснення зображень із втратами, VP8 все ще є високо конкурентоспроможним [48]. Нами у дослідженнях VP8 використовується як кодер стиснення з втратами для стиснення трьохканальних зображень RS. Нижче наведено опис того, як реалізувати двоетапний метод стиснення на VP8-кодері.

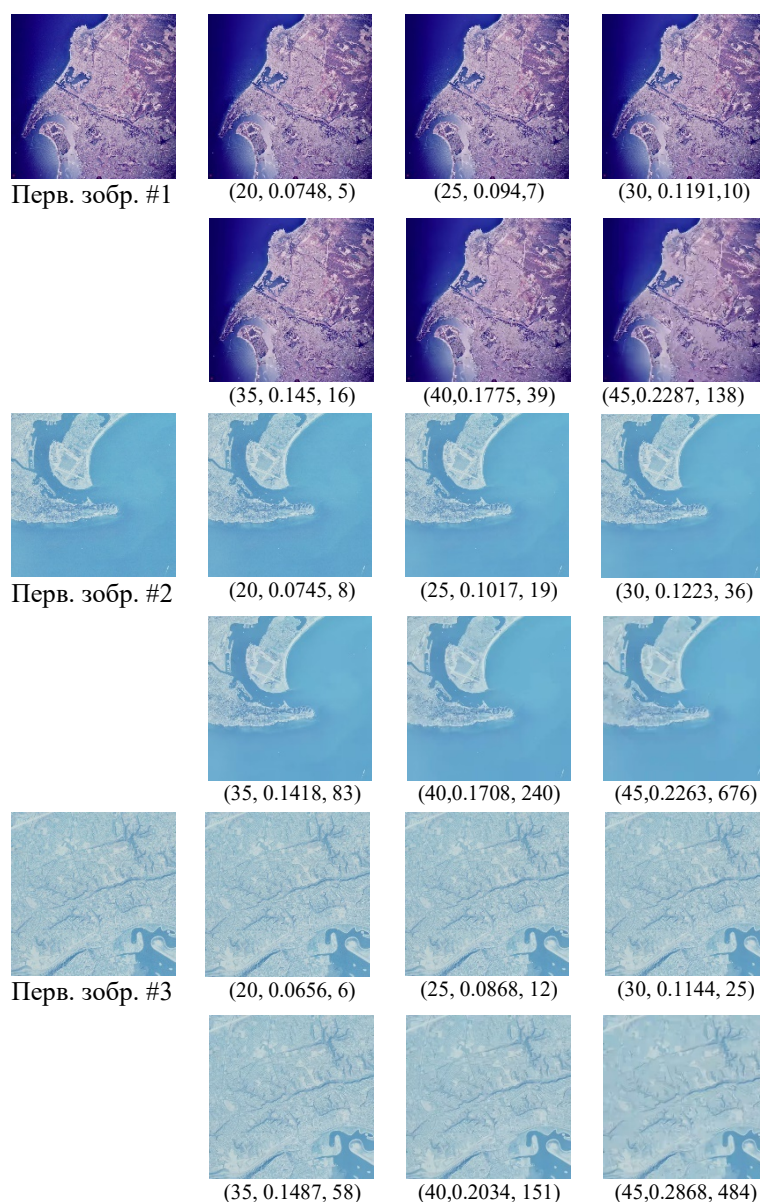
VP8-кодер заснований на техніці високоефективного кодування відео (High Efficiency Video Coding - HEVC), його запропонував Фабріс Беллар у вигляді відкритого коду [33]. У порівнянні з JPEG, VP8 забезпечує вищий коефіцієнт стиснення для тієї ж якості. Крім того, VP8 має такі властивості: 1) по-перше, він забезпечує декодер JavaScript, що збільшує доступність VP8 для веб-браузерів і портативних додатків завдяки його легкому коду; 2) по-друге, він є відкритим вихідним кодом, без оплати і патентів; 3) він підтримує різні формати представлення зображень, включаючи градації сірого, RGB, YCgCo та YCbCr; 4) VP8 підтримує стиснення без втрат і втрат. Всі ці функції роблять VP8 здатним замінити існуючі алгоритми.

В цьому кодері параметр Q розглядається як PCC, де  $Q \in \{1, \dots, 51\}$ . Більше значення Q призводить до вищого КС, що відповідає нижчій візуальній якості [33]. На відміну від PCC в AGU і SPIHT, параметр Q приймає тільки цілі значення. Кодер VP8 здатен приймати зображення форматів JPEG або PNG як вхідні дані, виконує стиснення і створює відповідне зображення у форматі VP8.

Основним завданням реалізації двоетапного методу на VP8-кодері є



отримання осередненої кривої КС/спотворення для розглянутої візуальної метрики якості (MDSI). Як вже зазначалося, ця крива базується на статистичних даних для певної кількості зображень. Ми обрали 12 триканальних зображень ДЗ як базовий набір зображень, усі вони мають розмір  $1024 \times 1024$  пікселів, ці зображення взяті з відомої бази USC-SIPi [49]. Стиснення та декомпресія виконуються для кожного зображення з різним значенням РСС. Використано всі можливі значення параметра  $Q$ . Деякі зображення, отримані в ході експериментів, показано на рисунку 7.



**Рисунок 7 – Три триканальні тестові зображення та дані про їх стиснення з різними значеннями параметру  $Q$**

Зліва показані три первинні зображення (без стиску), а потім для кожного з



них наведені по шість стиснених зображень для різних  $Q$  (20, 25, 30, 35, 40, 45), ці значення надані в дужках під кожним із зображенням першими в трійці чисел. Потім (другим у кожній трійці) надано значення MDPI. Нарешті, останнім в кожній трійці наведено значення КС. Зрозуміло, що візуально складно порівнювати такі маленькі зображення, але вони представлялись на екрані моніторів великого розміру з переключенням, щоб помітити різницю (якщо її було можливо помітити).

З аналізу даних на рисунку 7 можна дізнатися, що візуальна якість зменшується разом із збільшенням параметра  $Q$  - значення MDSI стає більше. Втім, для кожного зображення ця залежність є індивідуальною. Це доводить, що візуальна якість залежить не тільки від  $Q$ , але вона також пов'язана з зображенням, яке потрібно стиснути. Крім того, цей результат підтверджує наше твердження, наведене раніше. Якщо значення MDSI менше 0,15, спотворення в основному невидимі; вони стають помітними, коли MDSI збільшується до 0,2, наприклад, для тестового зображення №1 (45, 0,2287, 138), зображення №2 (45, 0,2263, 676) та №3(40, 0,2034, 151); спотворення очевидні, якщо значення MDSI більше 0,25, наприклад, для зображення #3(45, 0,2868, 484). Також можна помітити, що використання одного й того ж самого  $Q$  призводить до різних значень MDSI (найкращим прикладом є випадок  $Q=45$ ). З рисунку 7 можна також дізнатися, що КС збільшується, якщо  $Q$  стає більше. Для невидимих спотворень (MDSI близько 0,15) КС має порядок десятків і його значення значно залежать від складності зображення. Втім, для  $Q=45$  значення КС можуть мати порядок сотень.

Усі дані (12 зображень ДЗ, стиснені з 51 значенням параметра  $Q$ ) були записані у вигляді таблиці та отримані середні значення MDSI в результаті аналізу усіх зображень, що відповідають кожному значенню  $Q$ . На основі цих статистичних даних можна отримати криву КС/спотворення для кожного зображення, а також можна отримати осереднену криву КС-спотворення. Ці криві показані на рисунку 8. Нами ці криві розбито на чотири області для аналізу:

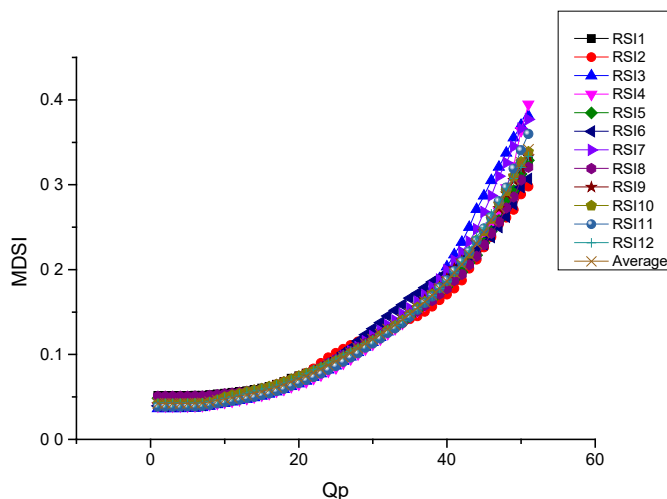
- а)  $Q \leq 3$ , візуальна якість постійна і дуже висока;
- б)  $4 \leq Q \leq 25$ , якість зображення починає повільно змінюватися, а значення MDSI для різних зображень дуже близькі;
- в)  $26 \leq Q \leq 40$ , якість зображень змінюється суттєвіше, але відхилення між показниками якості різних зображень все ще невеликі;
- г)  $41 \leq Q$  візуальна якість починає швидко знижуватися, а відхилення значень





MDSI для різних зображень стають очевидними. Отже, ідеальна робоча область  $Q$  становить  $[25; 40]$ , відповідна область значень MDSI становить  $[0,1; 0,25]$ .

Двоетапне стиснення може бути здійснене для BPG за допомогою осередненої кривої (наведеної на рисунку 8), але реалізація відрізняється від нашої попередньої схеми через наступні аспекти.



**Рисунок 8 – Індивідуальні та осереднена залежності MDSI від  $Q$**

По-перше, нам не потрібен вираз (3), що описує первинну лінійну апроксимацію, і будемо встановлювати початкове  $Q$  як найближче (або найближче менше) ціле число  $Q$ , отримуючи бажаний початковий MDSI відповідно до осередненої кривої. Іншими словами,  $Q_{init}$  дорівнює  $Q_{est}$ . По-друге, оскільки значення  $Q$  після корекції за допомогою рівняння (4), швидше за все, не ціле, то для отримання бажаного  $Q$  необхідна операція округлення.

Для оцінки ефективності запропонованого двоетапного методу для кодера BPG для триканальних зображень ДЗ необхідні тестові експерименти, оскільки модель осередненої кривої була отримана для базового набору зображень, який може бути не ідеальним. Отже, перш за все, ми застосували двоетапний метод стиснення для базового набору зображень (дванадцяти тестових зображень RSI #1...RSI #12), щоб отримати чотири типових бажаних значення для метрики MDSI. З огляду на аналіз, проведений вище, ці чотири типові значення встановлені як 0,1, 0,15, 0,2 та 0,25 відповідно. Статистичні дані наведені в таблиці 2.

Тут  $M_{des}$  позначає бажане значення MDSI,  $VAR_{fir}$  позначає СКП MDSI, отриману на першому кроці, а  $VAR_{sec}$  – це СКП MDSI, отримана на другому кроці. Для кращого розуміння даних нами також розраховано середнє значення





MDSI, яке надано для другого кроку, воно позначено як  $MEAN_{sec}$ .

**Таблиця 2 – Статистичні дані для базового набору**

$M_{des}$	$VAR_{fir}$	$VAR_{sec}$	$MEAN_{sec}$
0,1	$2,24 \times 10^{-5}$	$2,24 \times 10^{-6}$	0,1010
0,15	$4,31 \times 10^{-5}$	$6,73 \times 10^{-6}$	0,1495
0,2	$1,2 \times 10^{-4}$	$1,32 \times 10^{-5}$	0,1978
0,25	$3,32 \times 10^{-4}$	$1,85 \times 10^{-5}$	0,2466

Аналіз даних таблиці 2 показує, що для кожного бажаного значення СКП після другого кроку стиснення зменшилася приблизно на порядок. Це доводить, що двоетапна процедура добре працює в розглянутих умовах. Можна також побачити, що як  $VAR_{fir}$ , так і  $VAR_{sec}$  мають тенденцію до збільшення, якщо бажаний MDSI збільшується. Це означає, що завдання забезпечення потрібного MDSI є більш важливим для більших MDSI, як 0,2 або 0,25 (для  $M_{des}=0,1$  спотворення невидимі, і вони залишаються невидимими, якщо бажаний MDSI надається з помилкою близько 0,01). Середня помилка ( $|M_{des}-MEAN_{sec}|$ ) не перевищує 0,034, і ця середня помилка збільшується у міру зниження бажаної якості зображення, що схоже на тенденцію в наших попередніх роботах щодо інших кодерів [13-17].

Можливо, базовий набір містить недостатню кількість зображень і/або вони підібрані неправильно. Щоб перевірити це припущення, ми вибрали інші 12 зображень ДЗ як тестовий набір зображень (RSI #13...RSI #24). Як і базовий набір зображень, ці тестові зображення взято з бази зображень USC-SIPI [49], усі зображення мають розмір  $1024 \times 1024$  пікселів. Потім ми застосували метод двоетапного стиснення до цього тестового набору зображень, щоб перевірити правильність (універсальність) моделі кривої, отриманої з базового набору зображень. Статистичні дані наведені в таблиці 3.

У таблиці 3 для кожного бажаного MDSI CRG після другого кроку стиснення впала приблизно на порядок. Середня похибка не перевищує 0,035. Як бачимо, тенденції та значення подібні до наведених у таблиці 2. Це доводить, що базовий набір зображень вибрано правильно для отримання осередненої кривої КС-спотворення, і ця модель добре працює для інших триканальних зображень ДЗ.



Таблиця 3 – Статистичні дані для перевірного набору

$M_{des}$	$VAR_{fir}$	$VAR_{sec}$	$MEAN_{sec}$
0,1	$2,3 \times 10^{-5}$	$4,75 \times 10^{-6}$	0,1003
0,15	$5,48 \times 10^{-5}$	$6,33 \times 10^{-6}$	0,1490
0,2	$1,7 \times 10^{-4}$	$2,94 \times 10^{-6}$	0,1989
0,25	$3,01 \times 10^{-4}$	$2,04 \times 10^{-5}$	0,2465

Для детального аналізу отриманих результатів дані для 12 тестових зображень для  $MDSI_{des}=0,25$  наведені в таблиці 4. Тут  $Q_{init}$  – це параметр  $Q$ , який використовується для першого кроку стиснення. Він отриманий із осередненої кривої КС/спотворення і дорівнює 45 для всіх зображень. Як бачимо,  $MDSI$  після першого кроку різний для різних зображень, СКП дорівнює  $3,01 \times 10^{-4}$ , середнє значення 0,2448.  $Q_{des}$  позначає параметр  $Q$ , який використовується для другого кроку стиснення. Після другого кроку СКП становить  $2,04 \times 10^{-5}$ , а середнє значення дорівнює 0,2465. СКП для другого кроку становить 1/15 від СКП для першого кроку, а середнє також краще наближається до бажаного значення після другого кроку. Загалом точність докорінно підвищилася за рахунок другого кроку стиснення. Між тим, бувають випадки, коли  $MDSI$  після другого кроку такий самий, як і для першого кроку; наприклад, це відбувається для RSI №13. Це означає, що  $Q$  в таких випадках не потрібно виправляти, а другий крок стиснення не потрібен. У Таблиці 4, 5 із 12 зображень потребує лише одного кроку стиснення, щоб задовольнити вимогу точності, тоді як інші 7 зображень потребують другого кроку для підвищення точності. Для всіх проведених нами перевірок експериментів (для всіх 24 зображень і чотирьох бажаних  $MDSI$ ) 28,1% зображень потребує лише першого кроку двоетапного стиснення, щоб забезпечити бажаний візуальний ефект.

Щоб проаналізувати точність забезпечення бажаної візуальної якості для стиснення трьохканальних зображень дистанційного зондування на основі BPG із втратами, візьмемо три зображення як приклади, це RSI №13, №14 та №16. Бажана візуальна якість ( $MDSI$ ) дорівнює 0,25.

Розпаковані зображення для методу двоетапного стиснення розміщені в третьому стовпці на рисунку 8. Для бажаного значення  $MDSI$ , що дорівнює 0,25, початкове значення  $Q$  дорівнює 45, обчислені  $Q_{des}$  різні для різних зображень (вони відповідно дорівнюють 44, 45 та 47). Два зображення зліва від третього



Таблиця 4 – Детальні дані для перевірного набору зображень

Тест. зобр.	$Q_{init}$	$MDSI_{init}$	$Q_{des}$	$MDSI_{pro}$
RSI #13	45	0,2587	44	0,2427
RSI #14	45	0,2491	45	0,2491
RSI #15	45	0,2454	45	0,2454
RSI #16	45	0,2185	47	0,2447
RSI #17	45	0,2167	47	0,2394
RSI #18	45	0,2436	45	0,2436
RSI #19	45	0,2611	44	0,2445
RSI #20	45	0,2477	45	0,2477
RSI #21	45	0,2568	45	0,2568
RSI #22	45	0,2586	44	0,2443
RSI #23	45	0,2188	47	0,2507
RSI #24	45	0,2631	44	0,2489
СКП		$3,01 \times 10^{-4}$		$2,04 \times 10^{-5}$
MEAN		0,2448		0,2465

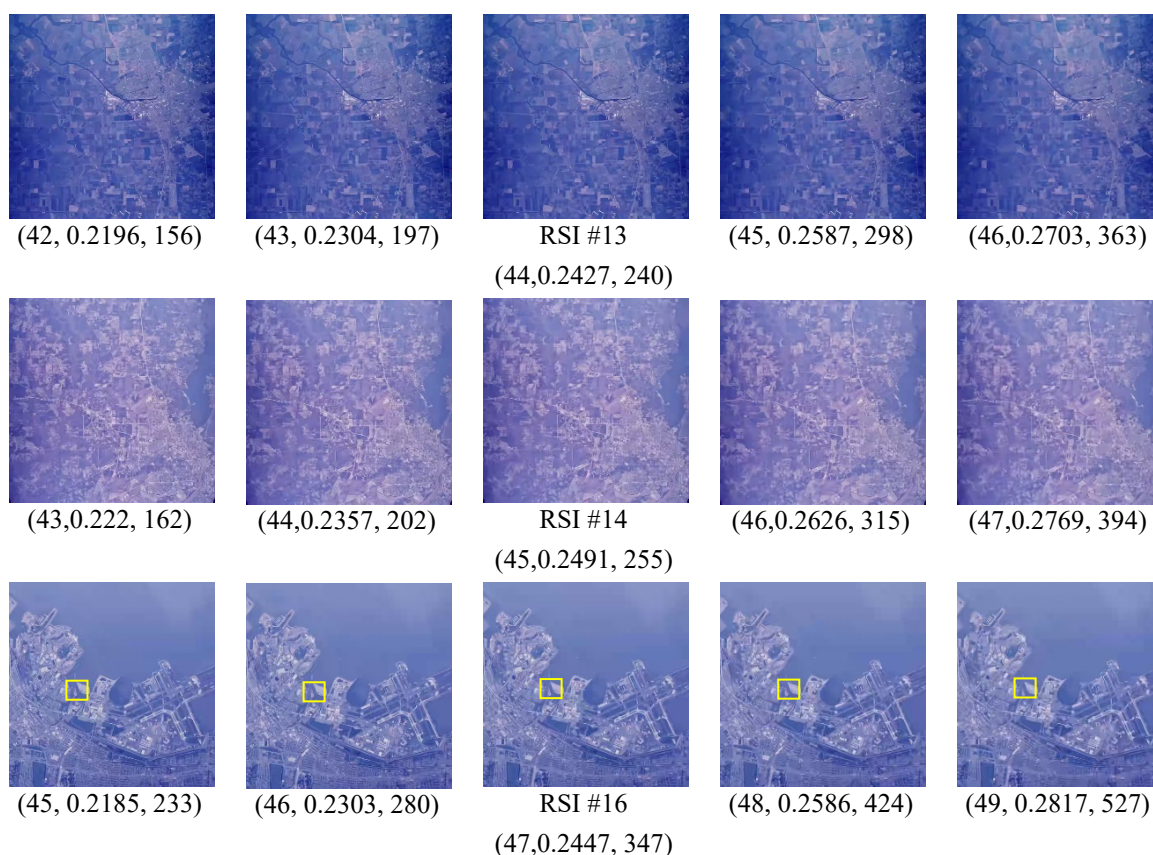
стовпця є зображеннями, коли  $Q$  встановлено як  $Q_{des-1}$  і  $Q_{des-2}$ , а два зображення праворуч є зображеннями, коли  $Q$  встановлено як  $Q_{des+1}$  і  $Q_{des+2}$ .

Для тестового зображення №14  $Q_{init}$  є відповідним значенням, незалежно від того, як його змінити (збільшити чи зменшити), похибка  $|MDSI_{pro}-MDSI_{des}|$  збільшується. Для тестового зображення №13  $Q$  після корекції становить 44 (початкове мінус 1), порівняно з чотирма іншими значеннями, стиснення за допомогою  $Q_{des}$  надає  $MDSI_{pro}$ , який є найближчим до  $MDSI_{des}$ . На відміну від цього, для тестового зображення #16 початкове значення  $Q$  коректується і становить 47 (плюс 2), стиснення з цим  $Q_{des}$  створює найменшу похибку між  $MDSI_{pro}$  та  $MDSI_{des}$ . Що стосується конкретних значень похибок, давайте візьмемо як приклад тестове зображення №16. Забезпечене  $MDSI$  дорівнює 0,2447, а помилка між  $MDSI_{pro}$  та  $MDSI_{des}$  становить 0,0053. На рисунку 9 показано п'ять розпакованих зображень №16. Зображення, стиснуті двома значеннями  $Q$ , що відрізняються на одиницю, здаються практично ідентичними, але якщо  $Q$  відрізняється на 2 або більше (див. дані для #16(45, 0,2185, 233) і RSI #16(49, 0,2817, 527)) порівняно з #16 (47, 0,2447, 347)), різницю легше відмітити. Це означає, що якщо різниця  $\Delta MDSI$  становить приблизно 0,015, важко спостерігати різницю в розпакованих зображеннях. Однак це стає помітним, якщо  $\Delta MDSI$  становить близько 0,03 (див. #16(46, 0,2303, 280) і #16(48, 0,2586, 424)). Отже, достатньо забезпечити похибки надання бажаного  $MDSI$  менше  $\approx 0,01$ . Тоді можна зробити висновок, що точність двоетапного методу для кодера



ВРГ є задовільною.

Підсумовуючи, потрібний другий крок чи ні, двоетапний метод може встановити відповідний параметр для кодера ВРГ у стисненні з втратами триканальних зображень RS. Незалежно від того, використовується пряма чи зворотна корекція (початкове значення Q збільшується чи зменшується), вона має позитивний вплив і в кінцевому підсумку забезпечує візуальну якість, яка є найближчою до бажаної. Значення КС, отримані для стиснення зображення з втратами в околиці порогу невидимості спотворення (дорівнює 0.25), значно вищі, ніж для стиснення без втрат.



**Рисунок 9 – Результати тестування двохкрокової процедури на трьох тестових зображеннях**

У цьому дослідженні нами запропоновано двоетапний (двохкроковий) алгоритм забезпечення бажаної візуальної якості для стиснення трьохканальних зображень дистанційного зондування на основі ВРГ. Для оцінки візуальної якості розпакованого зображення використано метрику MDSI. Основні результати полягають у наступному:

а) основний двохкроковий алгоритм модифіковано в тому сенсі, що він



враховує таку особливість BPG, що Q може бути лише цілим;

б) проаналізовано особливості метрики MDSI для оцінки якості трьохканальних зображень для 3000 зображень, надано межі області роботи для стиснення з втратами відповідно до цієї метрики. Пропонуються три рівні візуальної якості, які відповідають значенням MDSI: відмінна якість, хороша якість, середня та погана якість відповідно. Результати експерименту продемонстрували перевагу запропонованого алгоритму. Показано, що він дозволяє швидко стискати зображення із задовільними похибками забезпечення бажаної якості, що характеризується метрикою MDSI. Якщо параметр Q, розрахований для другого кроку, дорівнює початковому Q, це означає, що другий крок можна пропустити і процедуру прискорити. В іншому випадку потрібен другий крок, щоб підвищити точність із корекцією значення Q. Статистичні дані показують, що завдяки другому етапу точність значно покращується, а надана візуальна якість дуже близька до бажаної.

У майбутньому ми очікуємо, що стиснення з втратами на основі BPG може бути використане для забезпечення бажаних характеристик класифікації розпакованих зображень. Ми плануємо вивчити вплив візуальної якості, оціненої за допомогою MDSI, на точність класифікації, а також обговорити застосовність двоетапного алгоритму для BPG-кодера у застосуванні до класифікації багатоканальних зображень з високою роздільною здатністю.

### **5.3. Комбіновані метрики візуальної якості для об'єктивного оцінювання якості зображень із множинними спотвореннями**

В даний час зростає популярність і доступність відносно дешевих камер, а також електронних мобільних пристроїв, оснащених візуальними датчиками, що безсумнівно викликає динаміку зростання застосовності методів аналізу зображень і відео. Деякі очевидні приклади пов'язані з відеоспостереженням, моніторингом дорожнього руху, відеоінспекцією та діагностикою, відеонавігацією мобільних роботів або навіть автономних транспортних засобів. Дещо інші застосування стосуються неруйнівного контролю, спільної обробки даних від різних датчиків та багато інших, також пов'язаних із сучасними рішеннями Індустрії 4.0. Ще один фактор, що впливає на зростання популярності аналізу зображень, - це розробка деяких безкоштовних бібліотек, таких як





OpenCV, що дає можливість виконувати багато завдань у режимі реального часу, особливо з апаратною підтримкою, що забезпечується сучасними графічними процесорами (GPU).

Тим не менш, алгоритми машинного та комп'ютерного зору зазвичай використовують природні зображення, які можуть піддаватися різним спотворенням, що виникають не тільки під час їх отримання, але й викликані, наприклад, стисненням із втратами або наявністю помилок передачі. Це ситуація типова для сучасних електронних пристроїв, таких як фотоапарати, телефони та деякі інші гаджети, де зображення піддаються кільком нелінійним перетворенням перед записом. У такому випадку важливою є здатність виявляти такі спотворення та оцінювати загальну якість зображення.

За останні кілька років було запропоновано багато показників об'єктивної оцінки якості зображення (IQA), які можна розділити на три великі групи: з еталоном, які вимагають знання оригінального (первинного) зображення без будь-яких спотворень, методи без еталону, також відомі як «сліпі» метрики, та менш популярні підходи із обмеженим еталоном, які передбачають часткове знання еталонного зображення. Незважаючи на те, що методи без еталону є найбільш бажаними, їх універсальність і кореляція з суб'єктивними думками людей-спостерігачів, представлені як осереднені оцінки (MOS) або значення диференціальних MOS (DMOS) в базах даних IQA, як правило, значно нижчі в порівнянні з методами з еталоном. Більш детальний аналіз багатьох показників та їх порівняння для різних загальноприйнятих наборів даних, що містять еталонні та спотворені зображення разом із суб'єктивними оцінками якості можна знайти в деяких останніх дослідженнях [50–53].

Відомі численні спроби покращити кореляцію між метриками з еталоном і MOS (або DMOS). Один із способів зробити це — розробити так звані комбіновані метрики [50–53], які в той чи інший спосіб використовують кілька метрик (які ми називаємо елементарними). На практиці потрібні легко обчислювані метрики та простий спосіб їх комбінування, як і для 3D-друкованих поверхонь [54] або зображень дистанційного зондування [23]. Через це метою нашого дослідження, результати якого надані нижче у даному розділі, є отримати сімейство комбінованих метрик, які можна оптимізувати для такого застосування, як оцінювання якості зображень із множинними спотвореннями. Наскільки нам відомо, така оптимізація ще не проводилася для доступних баз даних, які містять лише зображення з численними спотвореннями.





розроблені комбіновані метрики [23, 54-60] стосуються зображень із частковими спотвореннями одного типу.

Найпоширеніші типи спотворення, до яких ідеальна метрика IQA має бути чутливою, стосуються артефактів розмиття, різних типів шуму та артефактів стиснення з втратами. Хоча в деяких наборах даних IQA, що містять спотворені зображення з одним типом спотворення, можна виділити більше 20 типів, наприклад, 24 типи в наборі даних TID2013 [61], включаючи спотворення, пов'язані з кольором, комбінації спотворень, надані в наборах даних IQA, обмежені лише кількома їх видами. Як правило, це комбінації розмиття, шуму, артефактів JPEG/JPEG2000 і зміни контрасту. Ці п'ять поширених типів викривлень були використані, наприклад, у базі даних MDID [62].

Враховуючи інтерференцію (одночасну присутність) окремих викривлень та їх вплив на якість зображення, корисність деяких комбінованих метрик, розроблених для зображень з окремими типами спотворень, для розвитку комбінованих показників, які сильно корелюють із суб'єктивною оцінкою якості зображень із множинними спотвореннями, не є очевидною і має бути перевірена експериментально.

Загалом ефективність комбінованих метрик якості залежить від таких факторів:

- кількість об'єднаних елементарних метрик;
- які показники поєднуються;
- як поєднуються показники;
- які зображення використовуються при тестуванні (оптимізації).

Отже, почнемо з швидкого пригадування сучасних елементарних метрик. Розробка сучасних візуальних показників якості, що замінила «класичні» MSE або PSNR, почалася фактично в 2002 році з ідеєю універсального індексу якості зображення (UQI) [63], а потім його вдосконалення, яке широко відоме як Structural Similarity (SSIM) [64], а також у його багатомасштабній версії (MS-SSIM) [65]. Деякі інші приклади, які також використовуються нижче, включають: SSIM з зваженим вмістом інформації (IW-SSIM) і IW-PSNR [66], Complex Wavelet SSIM (CW-SSIM) [67], Feature Similarity (FSIM) [68], індекс якості на основі локальної дисперсії (QILV) [69], а також кольорова версія SSIM (CSSIM), SSIM4 та його колірна версія CSSIM4 [70], що належать до групи метрик на основі SSIM з додатковою передбачуваністю блоків зображень.



Інший підхід, що походить з теорії інформації, передбачає використання статистики природної сцени у поєднанні з вимірюванням взаємної інформації між піддіапазонами в області вейвлетів, запропонованої Шейхом і Бовиком як метрика візуальної точності інформації (VIF) [71]. Його спрощена версія багатомасштабного піксельного домену (VIFp) вимагає менше обчислень, хоча не дозволяє аналізувати орієнтацію. Обидва методи засновані на попередній ідеї критерію вірності інформації (IFC) [72]. Метрика меншої обчислювальної складності, відома як подібність піддіапазонів ДКП (DSS) [73], використовує той факт, що статистика коефіцієнтів ДКП змінюється зі ступенем і типом спотворення зображення. Іншою мотивацією для його авторів була популярність 2D ДКП, оскільки багато методів кодування зображень і відео засновані на перетвореннях ДКП на основі блоків, зокрема, що походять із стандартів JPEG та MPEG.

Комбінація вейвлет-перетворення керованої піраміди та SSIM, відома як IQM2, була запропонована Думічем [74], де ядро з двома орієнтаціями було застосовано для досягнення найкращої продуктивності із збереженням низьких обчислювальних вимог. Інший підхід до перцептивного IQA був запропонований Ву [75], використовуючи внутрішній генеративний механізм (IGM), який приймає байєсівську модель передбачення та розкладає зображення на передбачені та невпорядковані частини. Передбачалося, що перша частина може бути оцінена за допомогою SSIM-подібних методів, тоді як невизначеність можна передбачити за допомогою PSNR. Обидві частини слід додатково нелінійно об'єднати, щоб отримати остаточну оцінку якості.

Чанг та ін. [76] запропонував метод, заснований на незалежній схожості ознак (IFS), що моделює властивості зорової системи людини (HVS), особливо корисний для прогнозування якості зображень із спотвореннями кольору. Через можливе використання часткової інформації з еталонного зображення (на основі використання Independent Component Analysis—ICA), цей метод також можна розглядати як приклад підходу з обмеженою інформацією про еталон. Інша метрика, заснована на HVS, відома як Perceptual Similarity (PSIM), була запропонована як чотириетапний метод [77] і частково перевірена за допомогою двох баз з множинними спотвореннями. Вона базується на використанні карт величин градієнта для обох порівнюваних зображень з подальшим обчисленням їх багатомасштабної подібності та вимірюванням деградації хроматичного каналу з наступним остаточним об'єднанням.



Крім того, автори метрики Sparse Feature Fidelity (SFF) [78] припускали перетворення зображень у розріджені представлення для виявлення розріджених ознак детектором ознак, навченим алгоритмом ІСА з використанням природних зразків зображень. Вони використовували компоненти подібності та кореляції яскравості для спільного моделювання візуальної уваги та зорового порога. Інша метрика, заснована на розріджених представленнях, відома як UNIQUE [79], використовувала підхід до навчання без вчителя. Цікаво, що на етапі попередньої обробки виконується вибір колірному простору (пропонується перетворення в модель YCbCr із заміною кольоровості Cb зеленим каналом) з подальшим вибором випадкових фрагментів, утворюючи вектор, що містить 64 елементи для кожного з трьох каналів, далі виконується нормалізація за допомогою віднімання середнього значення та операції вибілювання. Додаткове розширення шляхом аналізу вивчених ваг було запропоновано як метрика MS-UNIQUE [80]. Обидві метрики були навчені за допомогою випадково вибраних блоків із бази даних ImageNet. Також можливе подальше розширення такого підходу, заснованого на навчанні, зокрема з використанням підходів глибокого навчання [81, 82]; однак це все потребує відносно великого обсягу навчальних даних, доступних лише в окремих базах викривлених зображень.

Цікавою метрикою, що використовує подібність градієнта, подібність кольорів та об'єднання відхилень, є індекс середньої подібності відхилення (MDSI), яка вже розглядалася в другому розділі. Інша спроба використати подібність градієнта була запропонована Kсу [83], метрика відома як відхилення подібності за величиною градієнта (GMSD).

Райзенхофер та інші [84] запропонували використовувати вейвлет-декомпозицію Хаара для розробки іншої метрики перцептивної схожості на основі HVS, відомої як HaarPSI. Ця метрика заснована на використанні шести двовимірних вейвлет-фільтрів Хаара, що виділяють горизонтальні та вертикальні ребра на різних частотних шкалах, і може розглядатися як спрощення FSIM. Інший метод, заснований на функціях, відомий як RVSIM [85], використовує перетворення Рісса (подібно до RFSIM [86]) разом із візуальною контрастною чутливістю, тоді як метрика CVSSI [87] заснована на подібності контрасту та візуальної уваги, формуючи остаточну оцінку з використанням зважених стандартних відхилень локальної карти якості контрасту та глобальної карти візуальної уваги.

З огляду на тему нашого дослідження, наведений вище огляд елементарних



показників обмежується алгоритмами оцінювання з еталоном, які демонструють високу точність прогнозування для чотирьох розглянутих нижче наборів зображень, отриманих без будь-яких нелінійних функцій підгонки (наприклад, логістичних або поліноміальних). Хоча нещодавно було запропоновано кілька метрик, орієнтованих на оцінку якості для множинних спотворень, наприклад, з використанням виявлення градієнта [88], у деяких випадках їх коди недоступні або вони належать до групи «сліпих» методів, наприклад як метод, заснований на фазовій конгруентності [89]. Тому результати, представлені нами, сфокусовані на поєднанні більш відомих елементарних метрик з доступними кодами, спочатку розробленими для часткових спотворень.

На додаток до вищезгаданих показників, згадаємо деякі з методів IQA, які призвели до покращення ефективності, застосованої в комбінованих метриках: WSNR [89], PSNRHMA [90], VSNR [91], індекс візуальної стійкості (VSI) [92], Multiscale Contrast Similitude Deviation (MCSD) [93], спектральна залишкова подібність (SR-SIM) [94] та Wavelet Based Sharp Features (WASH) [95]. Деякі інші нещодавно запропоновані метрики, що використовуються в експериментах, спочатку були розроблені для оцінки якості зображень вмісту екрана, такі як SIQAD [96] і SCI\_GSS [97], а також для оцінки зміни контрастності зменшеної еталонної якості зображення (RIQMC) [98].

Розробка нових наборів баз для IQA є досить складним і трудомістким завданням, особливо якщо припустити проведення експериментів із залученням багатьох спостерігачів для відносно великої кількості спотворених зображень. Отже, серед багатьох наборів зображень для IQA лише деякі з них, наприклад, TID2013, що містять численні зображення, які піддаються кільком типам спотворення, можуть вважатися широко прийнятними спільнотою. На жаль, більшість баз зображень, розроблених в останні роки, не містять зображень з більш, ніж одним спотворенням, застосованим одночасно, і більшість метрик, розроблених і перевірених за допомогою таких наборів даних, оцінюють якість зображень з множинними спотвореннями з відносно низькою точністю.

Як стверджує Чандлер [51], однією з головних проблем у IQA зображень з множинними спотвореннями є той факт, що розроблені метрики повинні враховувати не тільки спільний вплив спотворень на зображення, а й вплив спотворень один на одного. Отже, беручи до уваги практичну корисність метрик, які могли б характеризувати візуальну якість зображень з множинними спотвореннями із можливою найвищою точністю, були розроблені деякі інші



набори даних, щоб заповнити цю прогалину в дослідженнях.

Перший з таких наборів даних, наданий Лабораторією інженерії зображень і відео (LIVE) Техаського університету в Остіні, іменованій LIVEMD [100], містить дві групи зображень з двома типами спотворень. Перша група має справу з розмиттям, а потім стисненням JPEG із втратами, тоді як друга група містить розмиті зображення через дефокусування, пошкоджене білим шумом для імітації шуму датчика. Кожна група містить 225 зображень, однак деякі з них фактично однократно спотворені, тому в експериментах, проведених у нашій роботі, було використано лише підмножину з 270 зображень з двома типами спотворень.

Інший набір даних, відомий як MDID13 [101], містить 12 еталонних зображень природних кольорів і 324 зображення, одночасно пошкоджених спотвореннями, які можуть мати місце під час отримання, стиснення та передачі зображень. Шість еталонних зображень стандартної чіткості (розміром  $768 \times 512$  пікселів) походять із бази даних Kodak, тоді як інші шість зображень високої чіткості ( $1280 \times 720$ ) такі ж, як у наборі даних LIVEMD. Тестові зображення містять суміш трьох чинників - розмиття, стиснення JPEG та шуму, що є доповненням до LIVEMD. Суб'єктивні оцінки були надані 25 недосвідченими спостерігачами, які використовували дві відстані перегляду через різні розміри зображення та метод одноразового стимулювання згідно з рекомендацією ITU-R BT.500-12.

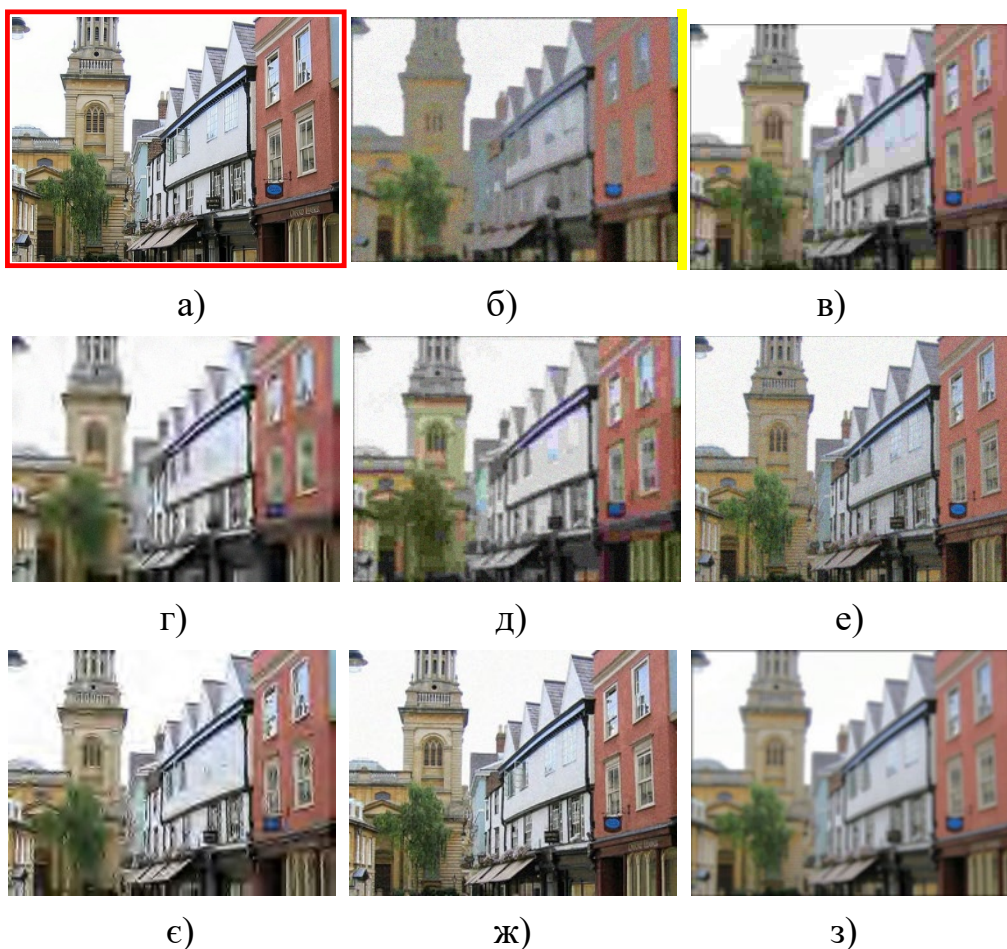
Третя база даних, яка використовується для перевірки запропонованого підходу, відома просто як MDID. Він містить 20 еталонних зображень (обрізаних до розміру  $512 \times 384$  пікселів без масштабування) і 1600 спотворених зображень. Зображення пошкоджені комбінаціями п'яти викривлень, а саме гаусовим шумом (GN), гаусовим розмиттям (GB), зміною контрасту (CC), JPEG та стисненням JPEG2000 із втратами. Кожне спотворене зображення було отримано з відповідного еталонного зображення із застосуванням випадкових типів і випадкових рівнів спотворення. Значення MOS надали 192 суб'єкти, які брали участь у суб'єктивному рейтингу. Зразки зображень із бази даних MDID, на які впливають різні комбінації спотворень з різними рівнями, представлені на рисунку 10 з еталонним зображенням, позначеним червоною рамкою.

Останній набір даних, розроблений в Лабораторії візуалізації та зору в Університеті Мілано-Бікокка, відомий як база даних IVL\_MD або MDIVL [102]. Він містить дві групи зображень: 400 зображень із шумами та спотвореннями JPEG, а також 350 зображень із розмиттям та спотвореннями JPEG разом із





відповідними значеннями MOS. Спотворені зображення, суб'єктивно оцінені 12 спостерігачами за методом одинарного стимулювання, були отримані з 10 еталонних зображень розміром 886 591 пікселів.



а) істинне зображення № 8; б) спотворене розмиттям за Гауссом (GB), зміною контрасту (CC), стисненням JPEG із втратами та гаусовим шумом (GN); в) спотворене CC, GB та JPEG; г) спотворене GB, JPEG2000 стисненням із втратами та GN; д) спотворені GB, JPEG та GN; е) спотворені CC, GB, JPEG2000 і GN; є) спотворений JPEG2000; ж) спотворені JPEG2000 і GN; з) спотворені GB, CC та JPEG2000.

**Рисунок 10 – Зразки зображень з бази даних MDID**

Існують також інші бази даних, що містять зображення з множинними спотвореннями, наприклад, LIVE у базі даних Wild Image Quality Challenge, що містить різноманітні автентичні спотворення зображень [103]. Однак ця база даних не надає еталонних зображень, а отже, не дозволяє розрахувати необхідні в нашому випадку показники якості з еталоном.





Порівнюючи чотири загальнодоступні бази зображень з множинними спотвореннями, безсумнівно, найбільш релевантною є база даних MDID не тільки через найбільшу кількість зображень і типів спотворення, але й з огляду на більшу кількість людей-спостерігачів, які беруть участь у експериментах. Тому експериментальні результати, отримані для цього набору даних, слід вважати найбільш важливими. З іншого боку, через більшу різноманітність спотворення та більшу кількість зображень очікувані значення кореляції нижчі, ніж для інших наборів даних.

Щоб забезпечити порівняння ефективності найкращих елементарних метрик для кожної з вищенаведених баз даних, були розраховані коефіцієнти лінійної кореляції Пірсона (PCC) між необробленими об'єктивними оцінками (тобто без будь-якої додаткової нелінійної підгонки) та суб'єктивними значеннями MOS/DMOS, що ілюструє точність метрик. Крім того, були розраховані коефіцієнти рангової кореляції Спірмена (SROCC) і Кендалла (KROCC), щоб проілюструвати монотонність поведінки кожної елементарної метрики.

Отримана ефективність для вибраних елементарних показників, у тому числі найкращих, представлена в таблиці 5, де три найкращі результати для кожного набору даних виділені жирним шрифтом. Як легко помітити, різні метрики демонструють найкращу ефективність для різних наборів даних, також відрізняючись точністю оцінювання, що вимірюється PCC, і монотонністю оцінювання, що характеризується ранговими кореляціями.

Хоча тут наведено не всі результати, отримані для елементарних метрик, для чотирьох розглянутих наборів даних було розраховано значення понад 50 з них. Крім того, були представлені результати кореляції, отримані для всіх баз даних, зважених за кількістю зображень у кожному з розглянутих наборів даних.

Таким чином, вагові коефіцієнти (до нормалізації) становлять 270 для LIVEMD (без урахування однієї спотвореної частини бази даних), 324 для MDID13, 1600 для MDID і 750 для MDIVL. Отже, найбільш «універсальними» елементарними метриками здаються VIF, DSS та IW-SSIM, які забезпечують найвищі кореляції, що є хорошою відправною точкою для розробки комбінованих метрик.



**Таблиця 5 – Значення коефіцієнтів кореляції елементарних метрик з MOS/DMOS для чотирьох баз зображень із множинними спотвореннями**

Метрика	База	Коефіцієнт кореляції		
		PCC	SROCC	KROCC
IW-PSNR	LIVEMD	0,5082	0,5111	0,3603
	MDID13	0,7649	0,7816	0,5697
	MDID	0,6859	0,6719	0,4846
	MDIVL	0,8303	0,8178	0,6229
	Зважена	0,6738	0,7064	0,5178
FSIM	LIVEMD	0,6954	0,6922	0,4803
	MDID13	0,5697	0,5818	0,3899
	MDID	0,8597	0,8873	0,7077
	MDIVL	0,7123	0,8589	0,6701
	Зважена	0,7743	0,8275	0,6415
VIF	LIVEMD	<b>0,7709</b>	<b>0,7588</b>	<b>0,5428</b>
	MDID13	0,8221	0,8447	0,6440
	MDID	<b>0,8873</b>	<b>0,9306</b>	<b>0,7714</b>
	MDIVL	<b>0,8568</b>	0,8378	0,6471
	Зважена	<b>0,8617</b>	<b>0,8817</b>	<b>0,7048</b>
DSS	LIVEMD	0,7070	<b>0,7439</b>	<b>0,5453</b>
	MDID13	0,7907	0,8078	0,5950
	MDID	0,8711	0,8658	0,6788
	MDIVL	0,8276	<b>0,8759</b>	<b>0,6910</b>
	Зважена	<b>0,8361</b>	<b>0,8508</b>	0,6604
IGM	LIVEMD	0,5527	0,6633	0,4606
	MDID13	0,8007	0,8239	0,6241
	MDID	0,8271	0,8548	0,6678
	MDIVL	0,7872	0,8637	0,6728
	Зважена	0,7889	0,8361	0,6453



## Продовження Таблиці 5

PSIM	LIVEMD	0,6883	0,6920	0,4800
	MDID13	<b>0,8325</b>	<b>0,8618</b>	<b>0,6630</b>
	MDID	0,8427	0,8733	0,6871
	MDIVL	0,7111	0,8427	0,6463
	Зважена	0,7939	0,8476	0,6550
HaarPSI	LIVEMD	0,6094	0,7155	0,5187
	MDID13	<b>0,8385</b>	0,8470	0,6425
	MDID	<b>0,8922</b>	0,8879	<b>0,7125</b>
	MDIVL	0,7936	0,8140	0,6212
	Зважена	0,8352	0,8487	0,6637
IW-SSIM	LIVEMD	<b>0,7398</b>	0,7377	0,5298
	MDID13	<b>0,8413</b>	<b>0,8551</b>	<b>0,6574</b>
	MDID	0,8634	<b>0,8911</b>	<b>0,7092</b>
	MDIVL	0,6955	0,8588	0,6708
	Зважена	0,8069	<b>0,8648</b>	<b>0,6773</b>
CSSIM4	LIVEMD	0,6664	0,6909	0,4850
	MDID13	0,8147	<b>0,8628</b>	<b>0,6665</b>
	MDID	0,5672	0,6639	0,4793
	MDIVL	0,6326	<b>0,9084</b>	<b>0,7320</b>
	Зважена	0,6202	0,7505	0,5648
VIFp	LIVEMD	0,7051	0,7142	0,5061
	MDID13	0,7361	0,7594	0,5561
	MDID	0,8184	0,8770	0,6978
	MDIVL	0,8000	0,7711	0,5721
	Зважена	0,7943	0,8221	0,6326
IQM2	LIVEMD	0,5087	0,6247	0,4305
	MDID13	0,7668	0,7806	0,5838
	MDID	0,8463	0,8530	0,6652
	MDIVL	<b>0,8681</b>	<b>0,8764</b>	<b>0,6891</b>
	Зважена	0,8121	0,8300	0,6408



## Продовження Таблиці 5

IFS	LIVEMD	0,6668	0,6729	0,4763
	MDID13	0,7132	0,7325	0,5305
	MDID	<b>0,9007</b>	<b>0,9070</b>	<b>0,7367</b>
	MDIVL	0,7032	0,8296	0,6388
	Зважена	0,8083	0,8466	<b>0,6652</b>
MDSI	LIVEMD	0,7059	0,6940	0,4842
	MDID13	0,6725	0,7024	0,4951
	MDID	0,8249	0,8360	0,6519
	MDIVL	0,8297	0,8376	0,6449
	Зважена	0,7985	0,8087	0,6175
RVSIM	LIVEMD	0,7139	0,7064	0,4835
	MDID13	0,6957	0,7253	0,5196
	MDID	0,8831	0,8835	0,7086
	MDIVL	<b>0,8626</b>	0,8517	0,6596
	Зважена	<b>0,8417</b>	0,8418	0,6547
CVSSI	LIVEMD	0,7059	0,7303	0,5266
	MDID13	0,7903	0,8065	0,5959
	MDID	0,8594	0,8638	0,6840
	MDIVL	0,8098	0,8540	0,6659
	Зважена	0,8239	0,8427	0,6552
UNIQUE	LIVEMD	0,7005	<b>0,7417</b>	<b>0,5357</b>
	MDID13	0,7004	0,8021	0,5983
	MDID	0,7691	0,7944	0,5888
	MDIVL	0,7678	0,7438	0,5498
	Зважена	0,7549	0,7775	0,5751

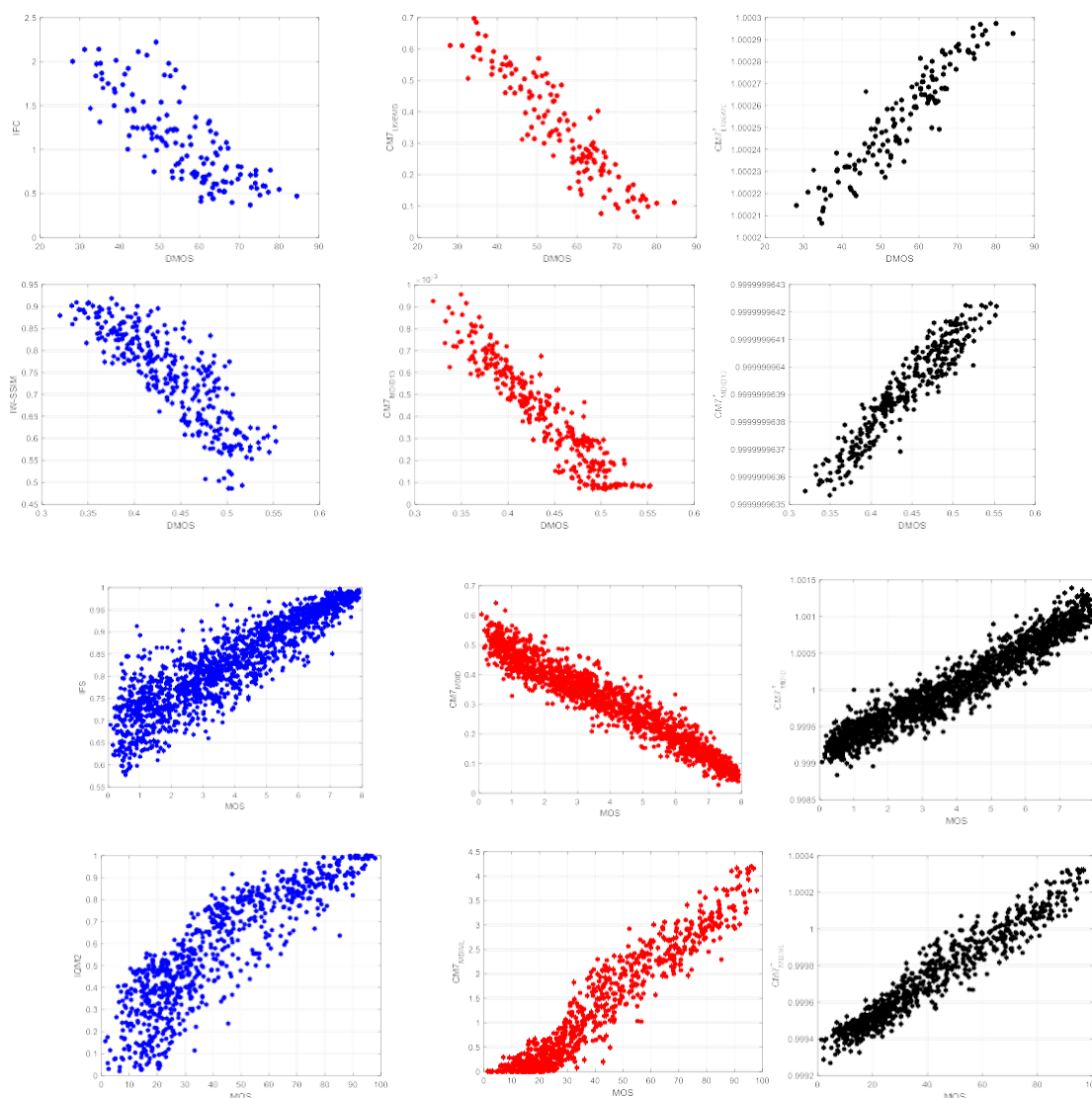


## Продовження Таблиці 5

SFF	LIVEMD	0,7205	0,7261	0,5197
	MDID13	0,7887	0,8005	0,5931
	MDID	0,8047	0,8396	0,6599
	MDIVL	0,7398	0,8535	0,6624
	Зважена	0,7787	0,8284	0,6403
MS-UNIQUE	LIVEMD	<b>0,7229</b>	0,7241	0,5120
	MDID13	0,7274	0,8316	0,6312
	MDID	0,7245	0,7423	0,5407
	MDIVL	0,7775	0,7550	0,5592
	Зважена	0,7382	0,7537	0,5528

В ідеалі метрика з еталоном повинна забезпечувати лінійну залежність між значеннями метрики та MOS. Менш строго, залежність між MOS і метрикою повинна бути монотонною (бажано, щоб більше значення метрики відповідало більшому MOS). Однак для багатьох існуючих елементарних метрик ці залежності далекі від ідеальних. Як приклади на рисунку 11 представлені діаграми розсіювання MOS від деяких елементарних метрик з еталоном для розглянутих баз даних (діаграми розсіювання в лівій колонці). Як бачимо, залежності можуть бути нелінійними (як для діаграми розсіювання IQM2 від MOS), різні метрики мають різні діапазони значень (багато показників варіюються в межах від 0 до 1, але не всі), деякі «викиди» (великі зміщення одних точок відносно більшості інших) також можуть статися. Ці властивості призводять до певних проблем при об'єднанні кількох елементарних показників у комбіновану метрику.

Ідея комбінованої метрики мотивується взаємодоповнювальними властивостями різних елементарних метрик, які можуть демонструвати «чутливість» до різного роду викривлень різного ступеня. Отже, було припущено, що їхня нелінійна комбінація може замінити необхідність нелінійної підгонки, запропонованого Групою експертів з якості відео (VQEG), щоб збільшити лінійну кореляцію між суб'єктивними та об'єктивними оцінками. Були зроблені деякі початкові спроби об'єднати метрики для одиночно спотворених зображень шляхом оптимізації вагових показників для добутку трьох метрик [54] з використанням бази даних TID2008, хоча під час подальших експериментів одна з метрик була замінена на FSIM, утворюючи комбіноване зображення. Індекс подібності (CISI) [55], який є зваженим добутком MS-SSIM, VIF та FSIM.



**Рисунок 11 – Скатерограми для “найкращих” елементарних метрик, отриманих для кожної з розглянутих баз зображень (ліва колонка), а також скатерограми для комбінованої метрики CM7 (центральна колонка) та метрики CM+7 (права колонка); зверху вниз дані для баз LIVEMD, MDID13, MDID та MDIVL**

У [56] було запропоновано мультиметричне злиття, засноване на регресійному підході для деяких старіших елементарних метрик з додатковою контекстно-залежною версією, що використовує підхід машинного навчання для автоматичного визначення контексту. Тим не менш, перевірку результатів проводили лише з використанням набору даних TID2008.

Інший підхід до комбінування метрик заснований на використанні генетичних алгоритмів для комбінації метрик [60], хоча моделюється як їх зважена сума замість їх добутку, що може обмежити можливість уникнення додаткової нелінійної підгонки. Отже, подібний підхід також використовувався





для зважених добутоків елементарних метрик [65], що призвело до подальших удосконалень.

Використання нейронних мереж для поєднання елементарних показників IQA було використано в роботі [61], де для навчання була використана випадково вибрана половина набору даних TID2013. Цей підхід використовував шість елементарних показників, що призвело до значного збільшення SROCC, обраного як критерій оптимізації. Проте, як і в інших випадках, комбіновані метрики використовувалися лише для оцінки спотворених зображень із одним типом спотворень. Крім того, потенційне застосування методів глибокого навчання потребуватиме розробки більших наборів навчальних даних, які містять також суб'єктивні оцінки якості для багаторазово спотворених зображень. Тому, як очікується, комбінація існуючих метрик із використанням відносно простої моделі буде ефективним рішенням також для зображень із множинними спотвореннями.

Для забезпечення простої форми комбінованої метрики, яка не потребує додаткової нелінійної регресії, наприклад, з використанням логістичної функції, у цій роботі спочатку була обрана стратегія, заснована на зваженому добутку елементарних метрик, а РСС був критерієм оптимізації. Хоча в деяких випадках монотонність може бути важливішою, ніж сама точність оцінювання, ми експериментально перевірили, що оптимізація показників зважування з використанням значень РСС як критерію також забезпечує високі значення SROCC. Під час експериментів виявилось, що результати, отримані в протилежному випадку, не завжди достатньо хороші. Іншою причиною використання РСС для необроблених оцінок без попереднього підбору нелінійності була гнучкість запропонованого підходу, що дає можливість контролювати всі ваги одночасно в одній процедурі оптимізації. Враховуючи різні динамічні діапазони елементарних показників, а також значення DMOS і MOS в кожному наборі даних, використання РСС не потребує додаткової нормалізації їх значень. Отже, припущена формула комбінованої метрики може бути виражена так

$$CM = \prod_{I=1}^N Q_i^{w_i}, \quad (5)$$

де  $N$  – кількість елементарних метрик, позначених як  $Q_i$ , а  $w_i$  – їх



експоненціальні ваги, отримані в результаті оптимізації, проведеної за допомогою функції `fminsearch` MATLAB.

Хоча застосування запропонованого методу комбінації показників дає обнадійливі результати, вибране об'єднання показників на основі їх зваженого добутку не завжди призводить до цілком задовільного результату. Отже, була також досліджена нова модель комбінування (злиття) на основі суми експоненціально зважених показників, де кожен компонент суми має додаткову вагу. Запропоновану формулу можна представити у вигляді:

$$CM^+ = \sum_{i=1}^N (a_i Q_i^w i), \quad (6)$$

де були введені додаткові коефіцієнти ваги  $a_i$ , щоб зробити комбінований показник ще більш гнучким і збільшити його кореляцію з суб'єктивними показниками якості, для найсучасніших наборів зображень із множинними спотвореннями.

Використовуючи вагові коефіцієнти  $a_i$  в рівнянні (6), враховуються різні діапазони варіації метрик (тобто виконується нормалізація). Використовуючи як коефіцієнти  $a$ , так і  $w$ , можна оптимізувати комбіновану метрику, тобто можна надати кращі значення PCC та/або SROCC порівняно з елементарними метриками, які використовуються як вхідні дані для комбінованої метрики.

Початкова перевірка корисності запропонованого підходу для оцінки якості зображень з еталоном для випадку множинних спотворень була здійснена в першу чергу для метрик, наведених у таблиці 5, з використанням чотирьох розглянутих наборів даних незалежно. Усі спочатку розглянуті метрики, що забезпечують значення PCC нижче нижньої межі, припущеної для всіх наборів даних, були виключені з початкових експериментів (тобто принаймні одна з умов має виконуватися для кожної метрики для включення в подальші експерименти). Значення цих меж для PCC: 0,7 для LIVEMD, 0,8 для MDID13, 0,85 для MDID і 0,8 для MDIVL. Відносно низька межа для набору даних LIVEMD спричинена видаленням окремо спотворених зображень з аналізу, що призводить до зменшення значень кореляції для цього набору даних. Тим не менш, у деяких випадках комбінації двох або трьох «гірших» показників можуть дати кращі результати в порівнянні з комбінацією одного з них з найефективнішим елементарним показником. Тому на другому етапі експериментів усі комбінації



двох і трьох метрик були перевірені для всіх наборів даних. Щоб обґрунтовано обмежити кількість можливих комбінацій, було обрано кілька «найкращих» комбінацій як основу для подальшого збільшення кількості показників.

Оптимізація експоненційних параметрів  $w_i$  для комбінованої метрики CM, а також множників  $a_i$  та експонент  $w_i$  для запропонованої формули CM+ була проведена за допомогою методу без похідних без обмежень на основі симплексного методу Нелдера–Міда, реалізованого у функції `fminsearch` MATLAB. Нарешті, усі множники  $a_i$  у запропонованій формулі CM+ були нормовані так, що  $\sum a_i = 1$ .

Оскільки «найкращі» комбінації двох, трьох і більше показників для окремих баз даних відрізняються, вони представлені в таблиці 6 окремо для кожного набору даних. Аналізуючи отримані результати, можна помітити, що для всіх наборів даних було досягнуто суттєве підвищення точності прогнозування навіть при використанні «найкращої» комбінації двох-трьох елементарних показників із використанням зваженого добутку метрик, позначених як CM. Використання більшої кількості додаткових елементарних метрик значно покращує отримані результати щодо PCC, а в деяких випадках може призвести до незначного зниження монотонності прогнозу (нижчі значення SROCC та KROCC).

**Таблиця 6 – Характеристики комбінованих метрик**

База	Набори	Коефіцієнти кореляції			Позначення
		PCC	SROCC	KROCC	
LIVEMD	IFC	0,7871	0,7891	0,5869	(елементарна)
	IW-SSIM, CSSIM	0,8637	0,8669	0,6741	CM2LIVEMD
	FSIM, IW-SSIM, SSIM4	0,8880	0,8853	0,7040	CM3LIVEMD
	FSIM, IW-SSIM, SSIM4, GMSD	0,8967	0,8900	0,7097	CM4LIVEMD
	FSIM, IW-SSIM, SSIM4, GMSD, CSSIM	0,9055	0,9037	0,7316	CM5LIVEMD



## Продовження Таблиці 6

	FSIM, IW-SSIM, SSIM4, GMSD, CSSIM, UNIQUE	0,9132	0,9107	0,7406	CM6LIVEMD
	FSIM, IW-SSIM, SSIM4, GMSD, CSSIM, UNIQUE, CSSIM4	0,9171	0,9135	0,7435	CM7LIVEMD
MDID13	IW-SSIM	0,8413	0,8551	0,6574	(елементарна)
	VSNR, CSSIM4	0,8930	0,9007	0,7159	CM2MDID13
	PSIM, VSNR, CSSIM4	0,9133	0,9171	0,7418	CM3MDID13
	PSIM, VSNR, CSSIM4, WSNR	0,9193	0,9214	0,7506	CM4MDID13
	PSIM, VSNR, CSSIM4, WSNR, RIQMC	0,9235	0,9261	0,7606	CM5MDID13
	PSIM, VSNR, CSSIM4, WSNR, RIQMC, CVSSI	0,9280	0,9304	0,7649	CM6MDID13
	PSIM, VSNR, CSSIM4, WSNR, RIQMC, SR-SIM, FSIM	0,9342	0,9370	0,7769	CM7MDID13
MDID	IFS	0,9007	0,9070	0,7367	(елементарна)
	IFC, MCSD	0,9456	0,9478	0,7999	CM2MDID
	IFC, MCSD, UQI	0,9520	0,9545	0,8132	CM3MDID
	IFC, MCSD, UQI, QILV	0,9542	0,9566	0,8173	CM4MDID
	IFC, MCSD, UQI, QILV, MS-UNIQUE	0,9559	0,9586	0,8215	CM5MDID
	IFC, MCSD, UQI, QILV, MS-UNIQUE, RVSIM	0,9579	0,9608	0,8259	CM6MDID
	IFC, MCSD, UQI, QILV, MS-UNIQUE, RVSIM, IW-SSIM	0,9587	0,9606	0,8261	CM7MDID



## Продовження Таблиці 6

MDIVL	IQM2	0,8681	0,8764	0,6891	(елементарна)
	SIQAD, CSSIM4	0,9400	0,9142	0,7431	CM2MDIVL
	QILV, SR-SIM, CSSIM4	0,9474	0,9291	0,7659	CM3MDIVL
	QILV, SR-SIM, CSSIM4, SIQAD	0,9502	0,9292	0,7675	CM4MDIVL
	QILV, SR-SIM, CSSIM4, SIQAD, CW-SSIM	0,9537	0,9410	0,7866	CM5MDIVL
	QILV, SR-SIM, CSSIM4, SIQAD, CW-SSIM, PSNRHMA	0,9553	0,9429	0,7901	CM6MDIVL
	QILV, SR-SIM, CSSIM4, SIQAD, CW-SSIM, PSNRHMA, VSI	0,9560	0,9441	0,7923	CM7MDIVL

Результати застосування запропонованої метрики CM+ на основі нормалізованої суми експоненціально зважених елементарних метрик представлені в таблиці 7, де вищі кореляції порівняно з відповідними метриками CM виділені жирним шрифтом. Як можна помітити, отримана ефективність запропонованих комбінованих показників краща для трьох наборів даних і трохи гірша для бази даних MDID. Додаткове порівняння лінійності досягнутої кореляції (без необхідності будь-якого додаткового нелінійного відображення) представлено на діаграмах розсіювання, показаних на рисунку 11.

Однак слід мати на увазі, що багато елементарних метрик мають різні властивості та різні динамічні діапазони, отже, тенденції, показані на різних графіках, можуть бути зворотними один до одного. Для деяких із цих метрик менші значення вказують на вищу якість, тоді як для деяких інших показників вірно навпаки. Оскільки максимальне абсолютне значення PCC розглядається як цільова функція, представлення діаграм розсіювання з використанням необроблених оцінок цих показників може представляти як «негативні», так і «позитивні» тенденції.





Таблиця 7 – Характеристики комбінованих метрик

База	Набори	Коефіцієнти кореляції			Позначення
		PCC	SROCC	KROCC	
LIVEMD	IFC	0,7871	0,7891	0,5869	(елементарна)
	IW-PSNR, SCI_GSS	0,8512	0,8498	0,6536	CM+2 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM	0,8732	0,8720	0,6844	CM+3 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM, SSIM4	<b>0,9075</b>	<b>0,9042</b>	<b>0,7359</b>	CM+4 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM, SSIM4, UNIQUE	<b>0,9118</b>	<b>0,9047</b>	<b>0,7390</b>	CM+5 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM, SSIM4, UNIQUE, IQM2	<b>0,9299</b>	<b>0,9231</b>	<b>0,7621</b>	CM+6 <sub>LIVEMD</sub>
	FSIM, IW-SSIM, SSIM, SSIM4, UNIQUE, IQM2, CVSSI	<b>0,9357</b>	<b>0,9302</b>	<b>0,7738</b>	CM+7 <sub>LIVEMD</sub>
MDID13	IW-SSIM	0,8413	0,8551	0,6574	(елементарна)
	VSNR, CSSIM4	<b>0,9013</b>	<b>0,9053</b>	<b>0,7253</b>	CM+2 <sub>MDID13</sub>
	VSNR, PSIM, MS-UNIQUE	<b>0,9228</b>	<b>0,9247</b>	<b>0,7577</b>	CM+3 <sub>MDID13</sub>
	VSNR, PSIM, MS-UNIQUE, WSNR	<b>0,9272</b>	<b>0,9260</b>	<b>0,7636</b>	CM+4 <sub>MDID13</sub>
	VSNR, PSIM, MS-UNIQUE, WSNR, SIQAD	<b>0,9329</b>	<b>0,9319</b>	<b>0,7727</b>	CM+5 <sub>MDID13</sub>



## Продовження Таблиці 7

	VSNR, PSIM, MS-UNIQUE, WSNR, SIQAD, QILV	<b>0,9372</b>	<b>0,9347</b>	<b>0,7742</b>	CM+6 <sub>MDID13</sub>
	VSNR, PSIM, MS-UNIQUE, WSNR, SIQAD, QILV, RFSIM	<b>0,9422</b>	<b>0,9423</b>	<b>0,7901</b>	CM+7 <sub>MDID13</sub>
MDID	IFS	0,9007	0,9070	0,7367	(елементарна)
	IFC, MCSD	0,9447	0,9459	0,7955	CM+2 <sub>MDID</sub>
	IFC, IFS, WASH	0,9517	0,9513	0,8029	CM+3 <sub>MDID</sub>
	IFC, IFS, WASH, VSI	0,9521	0,9534	0,8077	CM+4 <sub>MDID</sub>
	IFC, IFS, WASH, VSI, SSIM	0,9552	0,9569	0,8154	CM+5 <sub>MDID</sub>
	IFC, IFS, WASH, VSI, SSIM, IW-SSIM	0,9574	0,9581	0,8180	CM+6 <sub>MDID</sub>
	IFC, IFS, WASH, VSI, SSIM, IW-SSIM, MS-UNIQUE	0,9581	0,9594	0,8205	CM+7 <sub>MDID</sub>
MDIVL	IQM2	0,8681	0,8764	0,6891	(елементарна)
	SIQAD, CSSIM4	0,9381	0,9098	0,7372	CM+2 <sub>MDIVL</sub>
	DSS, QILV, SSIM4	<b>0,9510</b>	<b>0,9485</b>	<b>0,7975</b>	CM+3 <sub>MDIVL</sub>
	DSS, QILV, SSIM4, IW-PSNR	<b>0,9529</b>	<b>0,9500</b>	<b>0,8013</b>	CM+4 <sub>MDIVL</sub>
	DSS, QILV, SSIM4, IW-PSNR, CSSIM4	<b>0,9586</b>	<b>0,9581</b>	<b>0,8169</b>	CM+5 <sub>MDIVL</sub>
	DSS, QILV, SSIM4, IW-PSNR, CSSIM4, SIQAD	<b>0,9606</b>	<b>0,9575</b>	<b>0,8168</b>	CM+6 <sub>MDIVL</sub>
	DSS, QILV, SSIM4, IW-PSNR, CSSIM4, SIQAD, CW-SSIM	<b>0,9625</b>	<b>0,9608</b>	<b>0,8249</b>	CM+7 <sub>MDIVL</sub>

Це залежить від отриманих результатів оптимізації та елементарних метрик, які були використані в кінцевій комбінованій метриці. Оскільки в двох наборах даних значення DMOS були надані як суб'єктивні оцінки, тоді як винахідники двох інших наборів даних використовували значення MOS, вихідні значення —



різні для різних наборів даних — були використані нами та представлені у таблицях. Масштаб всіх отриманих комбінованих метрик залежить від вихідних оцінок окремих метрик, і отримані результати не були нормалізовані. Слід також зазначити, що високі значення DMOS зазвичай означають погану якість, тоді як високі значення MOS вказують на високу якість зображень.

Як можна помітити, результати для метрики CM7+, отримані для набору даних MDID2013, відрізняються помітно менше, ніж для трьох інших баз даних. Тим не менш, високолінійні зв'язки між суб'єктивними та об'єктивними оцінками якості досягаються в основному для запропонованих метрик CM+ для всіх розглянутих баз даних. Деякі відмінності в динамічних діапазонах комбінованих метрик, зокрема використання формул CM, є результатом використання різних типів метрик і різних ваг, отриманих після процедури оптимізації. Додаткове порівняння ефективності запропонованого підходу було проведено з використанням деяких інших комбінованих показників, раніше розроблених для однократно спотворених зображень, застосовуваних для наборів даних, що містять лише зображення із множинними спотвореннями. Отримані експериментальні результати для трьох таких наборів даних (MDID2013, MDID і MDIVL) представлені в таблиці 8. Оскільки чотири метрики на основі регресії (rSIM) були фактично розроблені як зважена сума окремих показників, була застосована додаткова нелінійна регресія з використанням логістичної функції з використанням коефіцієнтів, наведених у [59]. Як бачимо, наш підхід дає суттєво кращі результати, ніж підходи, запропоновані в [59, 60].

Графічна ілюстрація кореляції між «найкращою універсальною» комбінованою метрикою CM+7 та суб'єктивними оцінками для окремих наборів даних наведено на рисунку 12, де можна легко спостерігати найнижчу кореляцію для LIVEMD. Тим не менш, через найменшу кількість зображень цей набір даних можна вважати найменш значущим. Високолінійні співвідношення між суб'єктивною оцінкою та об'єктивною метрикою, досягнутою для трьох основних наборів даних (PCC = 0,9387 для MDID, PCC = 0,8911 для MDID13 і PCC = 0,9122 для MDIVL, відповідно, як показано на графіках на рисунку 12, підтверджують дієвість запропонованого підходу. Ці результати все ще кращі в порівнянні з результатами, отриманими для деяких альтернативних комбінованих метрик, представлених у таблиці 8.



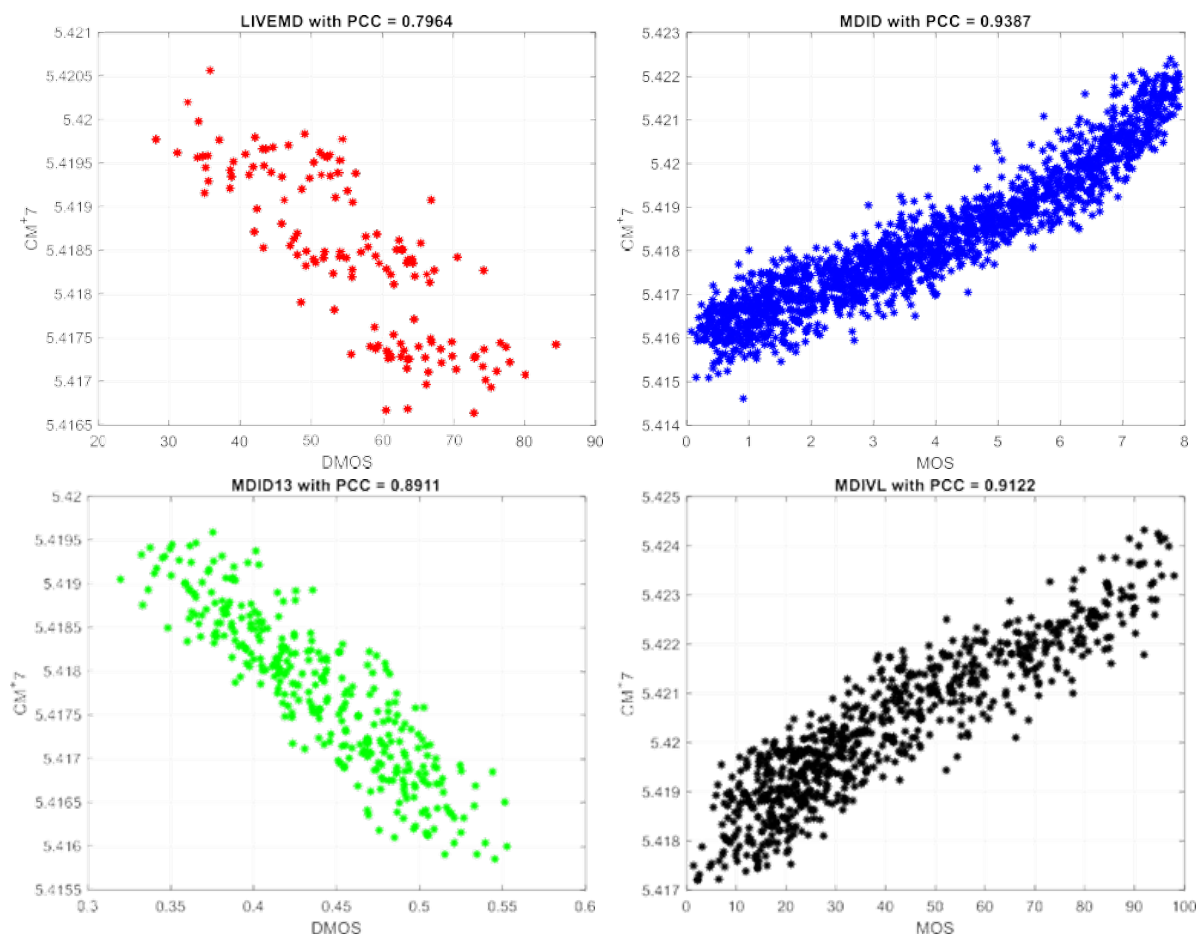
**Таблиця 8 – Порівняння результатів, отриманих для трьох основних наборів даних з використанням деяких комбінованих метрик, спочатку розроблених для зображень із одинарними спотвореннями з «найкращими» елементарними метриками та запропонованими метриками**

База	Метрика	Коефіцієнти кореляції		
		PCC	SROCC	KROCC
MDID13	IW-SSIM	0,8413	0,8551	0,6574
	CISI [54]	0,6882	0,6974	0,4894
	rSIM1 [59]	0,7416	0,7487	0,5454
	rSIM2 [59]	0,7438	0,7511	0,5529
	rSIM3 [59]	0,7469	0,7519	0,5471
	rSIM4 [59]	0,7464	0,7516	0,5476
	ESIM1 [60]	0,5807	0,5858	0,4030
	ESIM2 [60]	0,6666	0,6828	0,4794
	ESIM3 [60]	0,7034	0,7316	0,5250
	ESIM4 [60]	0,5773	0,5915	0,4015
	CM+7 (найкраща запропонована)	<b>0,9422</b>	<b>0,9423</b>	<b>0,7901</b>
MDID	IFS	0,9007	0,9070	0,7367
	CISI [54]	0,9045	0,9116	0,7427
	rSIM1 [59]	0,7443	0,7266	0,5344
	rSIM2 [59]	0,7429	0,7227	0,5320
	rSIM3 [59]	0,7453	0,7259	0,5342
	rSIM4 [59]	0,7442	0,7251	0,5334
	ESIM1 [60]	0,8704	0,8641	0,6805
	ESIM2 [60]	0,8780	0,8965	0,7247
	ESIM3 [60]	0,8977	0,9114	0,7448
	ESIM4 [60]	0,8752	0,8871	0,7089
	CM+7 (найкраща запропонована)	<b>0,9587</b>	<b>0,9606</b>	<b>0,8261</b>
MDIVL	IQM2	0,8681	0,8764	0,6891
	CISI [54]	0,8535	0,8599	0,6716
	rSIM1 [59]	0,8574	0,8734	0,6865
	rSIM2 [59]	0,7614	0,8089	0,5928
	rSIM3 [59]	0,8621	0,8651	0,6778
	rSIM4 [59]	0,8608	0,8653	0,6776
	ESIM1 [60]	0,7818	0,8319	0,6357
	ESIM2 [60]	0,8569	0,8452	0,6533
	ESIM3 [60]	0,7638	0,8477	0,6558
	ESIM4 [60]	0,7511	0,8583	0,6674
	CM+7 (найкраща запропонована)	<b>0,9625</b>	<b>0,9608</b>	<b>0,8249</b>



Вагові коефіцієнти, отримані для елементарних метрик, які мають різні властивості та різні динамічні діапазони, використані в  $SM+7$  згідно з формулою (6), наведені в таблиці 9.

Проведені експерименти підтвердили гіпотезу про те, що специфіка зображень з множинними спотвореннями потребує комбінації різних метрик, оскільки деякі із запропонованих раніше гібридних підходів призвели до гіршої ефективності навіть у порівнянні з «найкращими» елементарними метриками. Крім того, застосування комбінованої моделі, запропонованої нами, значно підвищує їх ефективність для більшості наборів даних. Застосування запропонованого підходу дає змогу підвищити як точність оцінки якості, виміряну PCC, так і монотонність оцінок, відображену обома ранговими кореляціями (SROCC та KROCC).



**Рисунок 12 – Діаграми розсіювання для «найкращої універсальної» елементарної метрики  $SM+$ , отриманої для кожного розглянутого набору даних, разом із значеннями PCC, отриманими для кожного набору даних незалежно. Суб'єктивні показники якості виражаються як MOS і DMOS, тоді як  $SM+$  позначає запроповану об'єктивну комбіновану метрику**





Оскільки запропонована модель комбінування метрик не є вимогливою до обчислень, її ефективність істотно не знижується, якщо припустити можливість паралельних розрахунків елементарних метрик. Єдиний виняток може бути пов'язаний з обмеженнями пам'яті, які можуть заважати паралельному обчисленню елементарних метрик для великих зображень. Вимоги до часу та пам'яті залежать від використовуваного обладнання та розміру зображення. Для паралельного обчислення метрик (наприклад, 7 метрик для 8 незалежних потоків) час обчислення кінцевої комбінованої метрики майже такий же, як і для «найповільнішої» елементарної метрики, що використовується.

**Таблиця 9 – Вагові коефіцієнти, отримані для елементарних показників, використаних у запропонованій «найкращій універсальній» метриці CM+7**

Елементарні метрики	Ваги	
	a	w
PSIM	3,7544	0,2552
IFC	0,2027	$7,9157 \times 10^{-4}$
GMSD	0,8024	28,8528
SIQAD	0,0678	2,5432
SVQI	$1,8587 \times 10^{-5}$	-0,0013
VIF	$-1,5179 \times 10^{-7}$	$7,0841 \times 10^{-4}$
FSIM	-0,0018	-0,0025

#### 5.4. Розробка метрик для оцінки якості панорамних зображень

Панорамні зображення – ще один тип зображень, що набувають популярності і для яких важливою є візуальна якість. Одним з найважливіших факторів, що впливають на кінцеву якість панорамних зображень, є відповідний вибір відмітних ознак зображення, які використовуються для відповідності тим самим регіонам, видимим на «сусідніх» складових зображеннях. Тим не менш, деякі додаткові операції пост-обробки, що виконуються після призначення, такі як змішування та інтерполяція, також можуть мати значний вплив на якість поєднаних зображень. Деякі очевидні приклади можуть бути пов'язані з різними умовами освітлення та змінами фону, видимими на складових зображеннях, які можуть викликати деякі легко помітні шви. Іншим фактором, пов'язаним із



геометричними спотвореннями, є вплив недосконалості об'єктива та надто малої кількості виявлених ключових точок, які використовуються для подальшого узгодження зображень, особливо для складових зображень із областями перекриття менше 15-20% площі зображення. Хоча деякі виправлення, наприклад калібрування, виправлення хроматичної аберації або віньєтування можна проводити за допомогою як безкоштовного, так і комерційного програмного забезпечення для зшивання зображень, навіть після остаточного змішування деякі недоліки все ще можуть бути помітні. Оскільки синхронне отримання складових зображень за допомогою кількох камер може бути проблемним у багатьох практичних застосуваннях, деякі проблеми також можуть виникнути для рухомих об'єктів, що призведе до розмиття в русі та фантомних артефактів.

Враховуючи сучасні застосування алгоритмів зшивання та реєстрації зображень, пов'язаних із використанням камер, встановлених на мобільних роботах, якість одержуваних панорамних зображень є дуже важливою через можливі помилки в зоровому контролі їх руху. У разі зниження якості зображення такі зображення можуть бути вилучені з аналізу, щоб запобігти їх впливу на керування роботом. Інший цікавий напрямок таких досліджень у мобільній робототехніці стосується злиття зображень, отриманих безпілотними літальними апаратами (БПЛА).

Приклади показані на рисунку 13. В роботі [4] розширення нещодавно запропонованої метрики SIQE до комбінованої метрики, запропонованої нами, дають змогу досягти значно вищої кореляції розроблених об'єктивних показників із суб'єктивними показниками якості з'єднаних зображень, які є у базі даних ISIQa.

Отримані результати є багатообіцяючими і підтверджують корисність комбінованих метрик також для автоматичної оцінки якості з'єднаних панорамних зображень. Як показано в дослідженні, застосування додаткових функцій на основі ентропії, що використовують локальну ентропію зображення та її дисперсію, є один із важливих елементів, що підвищують кореляцію зі значеннями MOS, оскільки їх видалення спричиняє найбільш значне зниження значень PLCC, SROCC та KROCC, очевидно, за винятком вихідної метрики SIQE.

Одним із потенційних напрямків подальших досліджень може бути застосування запропонованого підходу для оцінки якості методів зшивання



зображень, толерантних до паралакса, а також перевірка запропонованого підходу для деяких інших типів зображень і відеопослідовностей, що містять подібні види спотворення, також використовуючи деякі інші комбіновані моделі.



**Рисунок 13 – Приклад компонентних зображень та отриманих панорамних зображень**



## **Висновки**

В результаті виконання проекту розроблено комбіновані метрики для різних застосувань, як-то багатоканальні зображення дистанційного зондування, зображення контролю тривимірного друку, зображення із множинними спотвореннями, панорамні зображення тощо. Також для всіх цих видів зображення проведено дослідження, які дозволили виявити найбільш ефективні елементарні метрики. Це, в свою чергу, дозволило звужити коло елементарних метрик, які доцільно використовувати в комбінованих метриках та провести оптимізацію складу таких метрик та їх параметрів.

В результаті для більшості перелічених застосувань отримано РССтa SROCC на рівні 0,85 ... 0,95, що є задовільним для більшості з додатків.