

УДК 004.051

DOI: 10.31673/2412-9070.2020.065660

В. В. ГРЕБЕНЮК, аспірант;

О. А. ДІБРІВНИЙ, аспірант;

О. В. НЕГОДЕНКО, канд. техн. наук,

Державний університет телекомунікацій, Київ

ПОРІВНЯЛЬНИЙ АНАЛІЗ НЕРЕФЕРЕНТНИХ МЕТОДІВ ОЦІНЮВАННЯ ЯКОСТІ ВІДЕОМАТЕРІАЛУ

Проведено порівняльний аналіз функцій для оцінювання якості зображень за відсутності зразка: no-reference (NR) measure або методами NR-типу. Наявність NR-методів дуже актуально для оцінювання якості потокового відео (адже на боці отримувача відео немає еталона для порівняння якості) та оцінювання результатів перетворень, орієнтованих на поліпшення відео і вибір параметрів цих перетворень (яскравість зміни, стиснення динамічного діапазону яскравості, перетворення кольорового в півтон тощо). Досліджено шість типів експериментів із метою аналізу кореляції обчислюваних кількісних оцінок з візуальним оцінюванням якості тестованих відеофайлів. Три з них є принципово новими: порівняння відео після гамма-корекції і зміни контрасту з різними параметрами, а також розмиття, яке може бути наслідком розфокусування відеокамери. Також для порівняння було додано гібридний метод та метод повного референта. Експериментально показано, що жоден з досліджуваних нереперентних методів оцінювання якості зображення не є універсальним, а обчислена оцінка не може бути перетворена в якісну шкалу без урахування факторів, що впливають на спотворення якості зображення. Визначено, що більшість досліджених методів розраховує локальні оцінки для кожного кадру, а їхнє середнє арифметичне значення є оцінкою якості всього відеофайла. Якщо на відео домінують великі ділянки однорідної оцінки, методи такого типу можуть дати неправильні оцінки якості, які не збігаються з візуальним оцінюванням.

Ключові слова: оцінювання якості відео; NR-метод; RR-метод; FR-метод; гамма-корекція; контраст; розмите відео; шум.

ВСТУП

Розглядаючи два відео, людина може дійти висновку, що одне краще за інше або вони майже однакові. Вона може проранжувати три-п'ять коротких відео, візуально оцінивши їх якість, наприклад за п'ятибальною шкалою: дуже низьке, низьке, середнє, вище середнього, високе. Якщо потрібно візуально оцінити якість десятків або сотень відео, то це зробити вже важче, а проранжувати їх за рівнем якості практично неможливо.

Функції, що обчислюють кількісні оцінки відмінностей двох сигналів, відомі давно. Найбільш популярними з них є коефіцієнт кореляції, середньоквадратичне відхилення тощо. Цифрове зображення — це двовимірний сигнал, а відеоряд є послідовністю цифрових зображень, що чергуються одне за одним десятки разів за секунду. Схожість або відмінність відеозображень також можна оцінити подібними функціями, але вони не беруть до уваги кореляцію між сусідніми рядками і стовпцями. Це призвело до появи безлічі методів порівняння двох зображень. Зазвичай одне зображення вважається еталонним, а друге спотвореним, наприклад відновленим після стиснення. Метод дає можливість оцінити близькість другого зображення до першого. Безліч подібних методів і методик їх побудови детально описано в монографії [1]. Такі методи належать до класу методів порівняння з еталоном, а в англійській літературі їх називають full-reference (FR) measure.

На практиці еталонний відеоряд є не завжди (наприклад, онлайн-трансляція футбольного матчу, відеоконференція, трансляція відеогри тощо). У таких випадках (за відсутності еталона) для оцінювання якості відеосигналу застосовують так звані безеталонний метод, або no-reference (NR) measure. Називатимемо їх методами NR-типу. Також бувають гібридні методи, або reduced-reference (RR), це компроміс між двома попередніми методами. Вони використовують не повне зображення для порівняння, а, наприклад, службову інформацію відеофайла.

Слід зауважити, що функції даного типу не є метриками (як часто пишуть у літературі), оскільки не може бути обчислена «відстань» між двома об'єктами (у даному разі зображеннями) через наявність лише одного об'єкта.

Першу статтю, яка використовує термін no-reference, було опубліковано 2000 року [2], з того часу спостерігається нелінійне зростання публікацій на цю тему [3; 4]. Оскільки порівняльного розгляду методів досліджуваного типу практично не було, метою цієї роботи є порівняльний аналіз певної кількості безеталонних методів якості відеоматеріалів, дослідження кореляції отриманих кількісних оцінок із візуальним оцінюванням якості тестованих відео.

Якість зображення — це суб'єктивне поняття, тому висловити одним числом всі можливі аспекти суб'єктивного поняття якості досить складно. Перелічимо основні типи спотворень, які впли-

© В. В. Гребенюк, О. А. Дібрівний, О. В. Негоденко, 2020

вають на якість цифрового відео, сформованого оптичною системою:

- освітленість сцени, яку знімають;
- контрастність сцени, яку знімають;
- розмиття зображення в результаті руху камери або об'єкта зйомки;
- метод і ступінь стиснення зображення під час запису у файл;
- спосіб квантування зареєстрованого сигналу;
- глибина різкості.

Слід також зазначити, що некоректно порівнювати кількісні оцінки якості відеоматеріалів, які спотворено різними типами перетворень, наприклад розмите зображення з малококонтрастним. У таких випадках можна лише ділити зображення на два класи: «хороші» і «погані». Межі цих класів має встановлювати експерт у певній прикладній сфері, наприклад криміналістиці.

Під сценою будемо розуміти все, що потрапляє в кадр відеокамери. Передусім розглядатимемо спотворення, що вносяться алгоритмами квантування і стиснення, а саме блокування-артефакти (blocking artefacts). Недостатня глибина різкості також може впливати на якість зображення через розмиття деталей другого плану. Отже, серед аналізованих заходів будемо відшукувати такі, оцінки якості зображення яких збігаються з візуальним оцінюванням спостерігача під час перших чотирьох типах спотворень.

ОСНОВНА ЧАСТИНА

Методи для дослідження

У роботі було досліджено 11 методів. Для економії місця формули обчислення досліджуваних методів не подаються, але наводяться посилання на статті, що містять їх докладний математичний опис.

У результаті вивчення літератури було зроблено висновок, що багато методів, які використовуються для аналізу оцінки якості відеоматеріалу, ґрунтуються на звичайних NR-методах оцінювання якості зображень. Докладний огляд таких методів подано в роботі [7]. Коротко розглянемо, на чому базується обчислення досліджуваних методів. У переважній більшості методів, запропонованих у цій статті, обчислюється локальні оцінки в межах кожного кадру відеопотоку, сумуються, а потім загальну оцінку ділять на загальну кількість кадрів, визначаючи їх середнє арифметичне як фінальну оцінку. Також для порівняння було додано до списку один гібридний метод (RR — Reduced Reference) VQM та один повнореферентний метод PSNR.

Наведемо аббревіатури досліджуваних методів і опишемо величини, обчислювані в одній ділянці:

- BREN — квадрат різниці яскравості лівого і правого сусідів через один піксель [5];
- CONT — сума різниць яскравості центрального пікселя і його сусідів у ділянці [5];

- CURV — відношення яскравості пікселя до середньої яскравості в ділянці [5];

- STDI — середньоквадратичні відхилення яскравості в межах пікселя [5];

- CRET — зображення розмивається горизонтально і вертикально, обчислюються різниці яскравості сусідніх пікселів вихідного і розмитого зображення, метод передбачає максимальне сумарне розмиття за горизонталлю або вертикаллю щодо вихідного зображення [5];

- KURT — коефіцієнт ексцесу розподілу яскравості зображення [5];

- NIQE — витягує набір локальних функцій із зображення, а потім прирівнює до векторів багатовимірної моделі Гаусса (MVG) [6];

- BRISQUE — спочатку отримується природна статистика сцени (NSS), далі розраховуються вектори характеристик і на основі нейронної мережі намагається передбачити оцінку [7];

- PIQE — оцінює якість зображення на основі сприйняття людським оком [8];

- VQM — набір семи метрик відеопотоку [9];

- PSNR — розраховує різницю між шумом і оригінальним зображенням [10].

Експериментальні дослідження

Експериментальні дослідження виконувались із використанням одного оригінального відео та копій цього відео зі зменшенням якості відеоряду. Усього було шість відеофайлів. Розмір кадру становив 1920×1080 пікселів. Фрагменти (кадри) відеофайлів зображено на рис. 1. Перший кадр з оригінального зображення, а всі наступні кадри з меншою якістю. Зменшення якості відеоряду здійснювалося через збільшення ступеня стиснення зображень. Такий підхід дає можливість оцінити вплив сучасних методів стиснення відеофайлів під час завантаження їх на різні відеосервіси, на кшталт YouTube, GooglePhotos тощо. Як відомо, ці сервіси активно стискають матеріал для зменшення його фізичного розміру на диску. Також під час відеоконференцій або транслявання ігор у користувачів з обмеженою смугою пропускання інтернету може виникнути ситуація, коли якість відеопотоку зменшується. Передусім зменшується бітрейт відео (як наслідок збільшення компресії відеоряду). Оцінки обчислювались для напівтонових кадрів.



Рис. 1. Фрагменти відеоматеріалів для дослідження

Оцінювання якості після гамма-корекції відео. Було використано оцінку якості шести варіантів тестових відеозображень, отриманих гамма-корекцією з параметром γ , що змінювався від 0,5 до 3,5 з кроком 0,5 за формулою $\text{newIm}(i, j) = 255 \cdot (\text{Im}(i, j)/255)^\gamma$. Усього тестувалося $6 \cdot 6 = 36$ відеофайлів. Стоп-кадри з різних за якістю відео та вже після гамма-корекції ілюструє рис. 2.

Перший стоп-кадр — це оригінальне відео з гамма-корекцією 0,5, а останній — найгірша якість та гамма-корекція 3.

Це перетворення формує ряд зображень від світлого до темного з яскравістю, яка нелінійно змінюється. Найбільш якісні з візуального погляду зображення містяться в середині ряду (див. рис. 2).



Рис. 2. Фрагменти відеоматеріалу після гамма-корекції

Оцінювання якості після зміни контрасту зображення. Було сформовано шість варіантів кожного відеозображення з різними коефіцієнтами k контрасту від 0,5 до 3,5 з кроком 0,5, яскравість кожного пікселя (i, j) змінювалася відповідно до формули $\text{newIm}(i, j) = k \cdot (\text{Im}(i, j) - 127,5) + 127,5$ (рис. 3). Усього тестувалося $6 \cdot 6 = 36$ відеофайлів.



Рис. 3. Фрагменти відео після корекції контрасту

Оцінювання якості після лінійного приросту яскравості зображення. Для кожного відео виконувалося перетворення яскравості зображень типу $\text{newIm}(i, j) = \text{Im}(i, j) + c$, тобто до значення яскравості кожного пікселя (i, j) додавалася константа c (з урахуванням обмежень діапазону $[0, 255]$), що набуває значення від -100 до $+100$ з кроком 50, а також значення 75. Усього тестувалося $6 \cdot 6 = 36$ файлів (рис. 4).



Рис. 4. Фрагменти відео після корекції яскравості

Оцінювання якості після розмиття зображень. Для кожного відео виконувалося розмиття відеозображень стандартним фільтром Гаусса зі зміною значення сигми від 1,0 до 4,0 з кроком 0,75, а також оригінал без розмиття (рис. 5). Усього тестувалося $6 \cdot 6 = 36$ файлів.



Рис. 5. Фрагменти відео після розмиття

Оцінювання якості після додавання шуму. Для кожного відео виконувалося додавання шуму до відео стандартним білим шумом Гаусса зі зміною значення m від 0 до 0,5 з кроком 0,1 (рис. 6). Усього тестувалося $6 \cdot 6 = 36$ файлів.



Рис. 6. Фрагменти відео після додавання білого шуму

Оцінювання якості після додавання шуму типу сіль та перець (salt and pepper). Для кожного відео виконувалося додавання шуму до відео з використанням шуму типу сіль та перець зі зміною значення d від 0 до 0,5 з кроком 0,1 (рис. 7). Усього тестувалося $6 \cdot 6 = 36$ файлів.



Рис. 7. Фрагменти відео після додавання шуму типу сіль та перець

Результати експериментів розглянутих шести типів зібрано у таблиці. Чотири з них є принципово новими для тестування NR-методів, запропоновано три нових типи експериментів (№ 1–3).

Якщо кількісна оцінка методів збігалася з візуальною, записано знак «+», в іншому разі «-». Коли максимальне значення методів відповідало положому піку і було майже правильним, ставився знак «?». Нуль ставився, якщо для всіх варіантів зображень тесту значення методів були однаковими.

В останньому стовпці записано середній час оброблення відеофайла тривалістю 10 с та розмі-

Таблиця результатів порівняння 11 методів

Номер з/п	Метод	Тип експерименту						Час, с
		№ 1	№ 2	№ 3	№ 4	№ 5	№ 6	
1	BREN	? ++	+	+??+?	+++++	+++++	+++++	75
2	CONT	?	++?++	++?+?	+	++	++++	380
3	CURV	? ++	+	++?+?	+++++	+++++	++++	68
4	STDI	+++	+	+??+?	+++++	+++++	++++	81
5	CRET	?+		+?????			++	73
6	KURT	? ?+?	?		+++++	++++	++++	72
7	NIQE	? +?	+	+????	+++++	+++++	++++	90
8	BRISQUE	+++	+	+???	+++++	+++++	+++++	143
9	PIQE	? ?+?	+		+++++	+++++	++	87
10	VQM	+++++	+++++	+++++	+++++	+++++	+++++	120
11	PSNR	+	+++				+++++	62

CONT (380 с), VQM (120 с) та BRISQUE (143 с)

ром 1920 × 1080 пікселів на персональному Asus K401L. Характеристики комп'ютера: процесор Intel Core i3 5010U з тактовою частотою 2,1 ГГц, обсяг оперативної пам'яті 12 Гб, операційна система Windows 10.

Програми виконувалися в середовищі Matlab R2020b.

ВИСНОВКИ

У результаті виконаних порівняльних досліджень можна дійти таких висновків:

1. Виконано шість типів експериментів із метою аналізу кореляції обчислюваних кількісних оцінок із візуальним оцінюванням якості тестових відеофайлів. Три з них є принципово новими: порівняння відео після гамма-корекції і зміни контрасту з різними параметрами, а також на-вмисна зміна яскравості зображення.

2. Тест № 1 виявився досить складним: тільки чотири заходи (NIQE, BRISQUE, KURT та VQM) дали майже половину правильних або близьких до правильних відповідей.

3. Практично всі NR-методи (крім VQM, але це гібридний метод) на тестових відео показали поганий результат у тесті № 2, визначаючи зображення з максимальним коефіцієнтом контрасту як найкраще. Водночас у тесті № 4 (розмиття) оцінки практично всіх методів збіглися з візуальним оцінюванням якості.

4. Загалом найгіршим виявився метод PSNR, який погано спрацював на всіх тестах, часто виставляючи оцінку для спотворених відео вищу, ніж для оригіналів.

5. Цікаву особливість було виявлено в методі CONT. Цей метод показав неправильні результати на всіх тестах, крім самого складного тесту № 2, де у нього була найбільша кількість правильних оцінок.

6. Усі заходи обчислюються швидко (менш ніж за 2 хв на відео 1920 × 1080 пікселів), за винят-

ком трьох: CONT (380 с), VQM (120 с) та BRISQUE (143 с).

7. У середньому не повну, але задовільну кореляцію з візуальним оцінюванням якості показали такі заходи: NIQE, BRISQUE, PIQE.

8. Тест № 2 продемонстрував, що для більшості методів найбільш якісним є максимальне контрастний варіант зображення, тобто практично чорно-білий. Отже, усі досліджені методи належать до класу контекстно-незалежних і не можуть повністю збігатися з візуальним оцінюванням якості зображень (не стосується методів RR та FR).

9. Експериментально показано, що жоден із тестованих методів оцінювання якості зображення не є універсальним, а обчислена оцінка не може бути перетворена в якісну шкалу без урахування факторів, що впливають на спотворення якості зображення (не стосується методів RR та FR).

10. Слід розробляти методи оцінювання якості, що враховують контекст змісту і формують оцінки, відмінні від середнього значення безлічі локальних оцінок.

Список використаної літератури

1. Старовойтов В. В. Локальные геометрические методы цифровой обработки и анализа изображений. Минск: Ин-т техн. кибернетики НАН Беларуси, 1997.

2. Impairment metrics for digital video and their role in objective quality assessment / J. Caviedes [et al.] // Visual Communications and Image Processing, Perth, Australia, 30 May 2000. P. 791–800.

3. Wang Z., Bovik A. C. Modern Image Quality Assessment // Synthesis Lectures on Image, Video, & Multimedia Processing, Morgan & Claypool, San Rafael, Calif, USA, 2006.

4. Muijs R., Kirenko I. A no-reference blocking artifact measure for adaptive video processing // Proc. of the 13th European Signal Processing Conference (EUSIPCO'05), Antalya, Turkey, September 2005.

5. *Pertuz S., Puig D., Garcia M. A. Analysis of focus measure operators for shape-from-focus // Pattern Recognition, 2013. Vol. 46. № 5. P. 1415–1432.*

6. *Wang Z., Bovik A. C., Evans B. L. Blind measurement of blocking artifacts in images // Proc. IEEE Int'l. Conf. Image Process., 2000. P. 981–984.*

7. *Learn OpenCV — Image Quality Assessment: BRISQUE. URL:*

<https://www.learnopencv.com/image-quality-assessment-brisque/> (Дата звернення 10.01.2021)

8. *MathWorks. URL:*

<https://www.mathworks.com/help/images/ref/piqe.html> (Дата звернення 10.01.2021)

9. *National Telecommunications and Information Administration (NTIA). URL:*

<https://www.its.bldrdoc.gov/resources/video-quality-research/standards/objective-models.aspx> (Дата звернення 10.01.2021)

10. *Thomos N., Boulgouris N. V., Strintzis M. G. Optimized Transmission of JPEG2000 Streams Over Wireless Channels // IEEE Transactions on Image Processing, 15. 2006.*

V. V. Grebeniuk, O. A. Dibrivniy, E. V. Nehodenko

СРАВНИТЕЛЬНЫЙ АНАЛИЗ НЕРЕФЕРЕНТНЫХ МЕТОДОВ ОЦЕНКИ КАЧЕСТВА ВИДЕОМАТЕРИАЛА

Проведен сравнительный анализ функций для оценки качества изображений при отсутствии образца: *no-reference (NR) measure* или методами *NR*-типа. Наличие *NR*-методов очень актуально для оценки качества потокового видео (ведь на стороне получателя видео нет эталона для сравнения качества) и оценки результатов преобразований, ориентированных на улучшение видео и выбор параметров этих преобразований (яркость изменения, сжатие динамического диапазона яркости, преобразование цветного в полутон и другие). Исследовано шесть типов экспериментов с целью анализа корреляции вычислительных количественных оценок с визуальным оцениванием качества тестируемых видеофайлов. Три из них являются принципиально новыми: сравнение видео после гамма-коррекции и изменения контраста с различными параметрами, а также размытие, которое может быть следствием расфокусировки камеры. Также для сравнения были добавлены гибридный метод и метод полного референта. Экспериментально показано, что ни один из исследуемых нереферентных методов оценки качества изображения не является универсальным, а исчисленная оценка не может быть преобразована в качественную шкалу без учета факторов, влияющих на искажение качества изображения. Показано, что большинство исследованных методов вычисляет локальные оценки для каждого кадра, а их среднее арифметическое значение является оценкой качества всего видеофайла. Если на видео доминируют крупные области однородной оценки, методы такого типа могут дать неправильные оценки качества, которые не совпадают с визуальными оценками.

Ключевые слова: оценка качества видео; *NR*-метод; *RR*-метод; *FR*-метод; гамма-коррекция; контраст; размытое видео; шум.

V. Grebeniuk, O. Dibrivniy, O. Nehodenko

A COMPARATIVE ANALYSIS OF NON-REFERENCE ASSESSMENT METHODS

A comparative analysis of functions to assess image quality in the absence of a sample: *no-reference (NR) measure* or *NR-type methods*. The availability of *NR*-methods is very important for assessing the quality of streaming video such as television, game streaming, online conferences, web-chatting, etc. (because on the side of the recipient of the video there is no standard for quality comparison) and assessing the results of transformations aimed at improving video, and choosing the parameters of these transformations (brightness change, semitone and others). The human visual system (HVS) is able to visually assessing video quality, but if required to visually assess the quality of dozens or hundreds of videos or ranking them by quality level it will be needed a huge amount of time. Six types of experiments were performed to analyze the correlation of calculated quantitative estimates with visual assessments of the quality of the tested video files. Three of them are fundamentally new: comparing video after gamma correction and changing the contrast with different parameters, as well as blurring, which may be the result of defocusing the camcorder. A hybrid method (or reduced-reference (*RR*) measure) and a full-reference (*FR*) measure or *FR*-type method were also added for comparison. It has been experimentally shown that none of the studied non-reference methods of image quality assessment is universal, and the calculated assessment cannot be converted into a quality scale without taking into account the factors influencing the distortion of image quality. Moreover, all *NR*-type methods could not cope with the experiment of changing the contrast, believing that the best result is the most contrasting image but the original. Instead, the reference methods showed an excellent result (except one, which showed partial ineffectiveness). Also, it has been shown performance comparison between methods. It is shown that most of the studied methods calculate local estimates for each frame, and their arithmetic mean value is an estimate of the quality of the entire video file. If the video is dominated by large areas of uniform evaluation, methods of this type may give incorrect quality evaluations that do not coincide with the visual evaluations.

Keywords: video quality assessment; *NR* method; *RR* method; *FR* method; gamma correction; contrast; blurred video; noise.