

SPÓJNA CZASOWO OBIEKTYWNA MIARA JAKOŚCI DLA WIZJI WSZECHOGARNIAJĄCEJ
TEMPORALLY-CONSISTENT OBJECTIVE QUALITY METRIC FOR IMMERSIVE VIDEO

Jakub Stankowski, Bartosz Sojka, Tomasz Grajek, Adrian Dziembowski

Instytut Telekomunikacji Multimedialnej, Politechnika Poznańska, Poznań
{jakub.stankowski, tomasz.grajak, adrian.dziembowski}@put.poznan.pl

Streszczenie: W niniejszym artykule przedstawiono efektywne rozwinięcie metryki IV-PSNR, służącej do obiektywnego pomiaru jakości dla wizji wszechogarniającej. Zaproponowany algorytm – IV-PSNR_{YUVM} – uwzględnia analizę ruchu w dziedzinie czasu. Analiza ta możliwa jest dzięki dodaniu wektorów ruchu jako czwartej składowej dla każdego piksela obrazu. Przeprowadzone eksperymenty pokazują, że zaproponowana modyfikacja znacząco zwiększa korelację między obiektywną i subiektywną jakością wizji wszechogarniającej.

Abstract: In this paper, an efficient enhancement to the IV-PSNR metric is introduced. The proposal – IV-PSNR_{YUVM} – incorporates a motion analysis in the temporal domain by adding motion vectors as a fourth component to video (Y, CB, and CR). Performed experiments demonstrate that this modification significantly improves the correlation between objective and subjective quality assessments in immersive video.

Słowa kluczowe: wizja wszechogarniająca, ocena jakości wizji, obiektywny pomiar jakości.

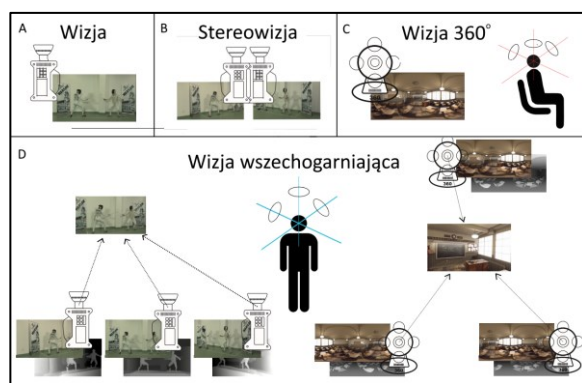
Keywords: immersive video, video quality evaluation, objective quality assessment.

1. WSTĘP

Wizja wszechogarniająca, zwana także immersyjną, to taka, która pozwala „zanurzyć się” widzowi w prezentowanym świecie. To zanurzenie może dotyczyć treści zarówno naturalnych (zarejestrowanych odpowiednimi zestawami kamer), jak i wygenerowanych komputerowo [3], [5]. Ciągłe prowadzone są prace nad wieloma technologiami związanymi z wizją wszechogarniającą, z których wszystkie prezentują różne poziomy/stopnie immersji (rys. 1). Niektóre są bardzo ograniczone, np. pozwalają użytkownikowi jedynie zmieniać kierunek widzenia [9], podczas gdy inne są dość zaawansowane, np. pozwalają użytkownikowi swobodnie się poruszać [12].

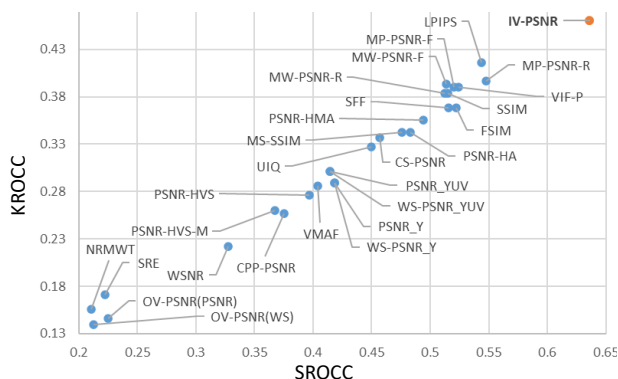
Pomimo iż systemy wizji wszechogarniającej (ang. *immersive video*) [1], [15] są o wiele bardziej zaawansowane niż tradycyjne systemy wizji dwuwymiarowej, ich wspólny cel pozostaje niezmienny: zapewnienie jak najlepszej jakości wrażeń (ang. *QoE, Quality of Experience*) poprzez maksymalizację subiektywnej satysfakcji użytkowników. Niestety, przeprowadzenie kompleksowych

testów subiektywnych dla oceny jakości jest pracochłonnym i czasochłonnym przedsięwzięciem [4], co czyni je wysoce niepraktycznymi. Dużo praktyczniejszym rozwiązaniem jest ocena jakości w sposób obiektywny.



Rys. 1. Rodzaje systemów wizyjnych. Rysunek z [5].

Dziedzina obiektywnego pomiaru jakości jest obszerna, szczególnie w obrębie przetwarzania obrazów i wizji. W wielu scenariuszach, takich jak tradycyjna wizja dwuwymiarowa, najnowsze metryki skutecznie symulują subiektywne postrzeganie jakości wizji, np. PSNR, SSIM [14] lub VMAF [11]. Jednakże, nietypowa charakterystyka wizji wszechogarniającej, obejmująca przetrzymywanie danych zarejestrowanych przez kilka kamer [12] czyni ją znacznie trudniejszą w ocenie, a tradycyjne metryki oceny jakości obiektywnej nie radzą sobie w jej kontekście wystarczająco skutecznie.

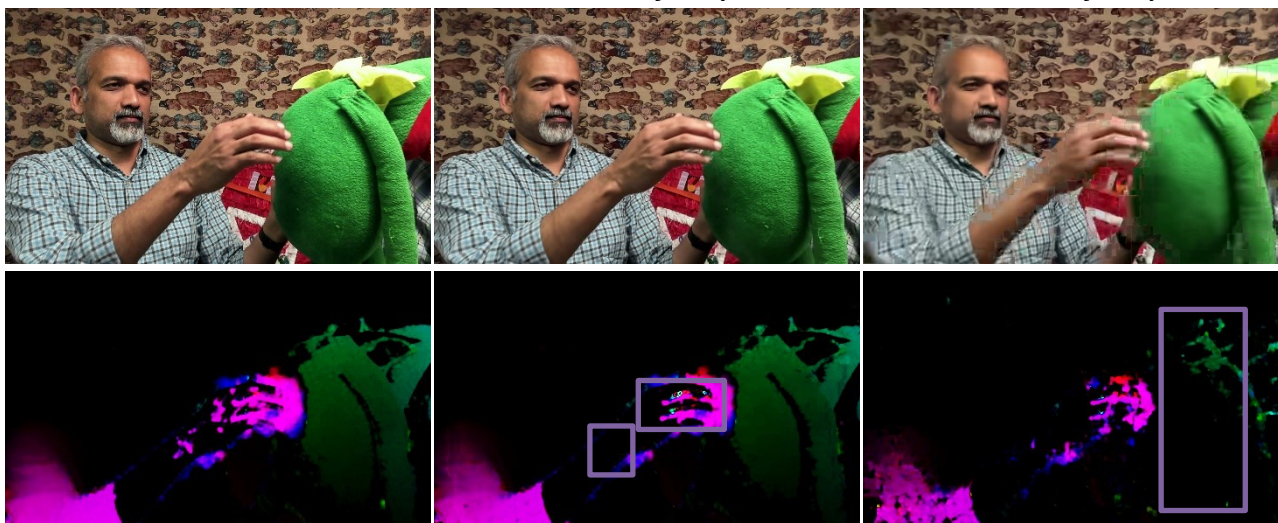


Rys. 2. Korelacja (mierzona z użyciem metryk SROCC i KROCC [10]) pomiędzy subiektywną oceną jakości a jakością mierzoną obiektywnie, z użyciem metryk opisanych w literaturze. Rysunek z [4].

(A) Widok wejściowy

(B) Widok stworzony (wyrenderowany) na podstawie nieskompresowanych widoków wejściowych

(C) Widok stworzony (wyrenderowany) na podstawie silnie skompresowanych widoków wejściowych



Rys. 3. Przykładowe pole ruchu wyznaczone dla sekwencji oryginalnej (A) oraz zniekształconej (B i C).

Aby uwzględnić zniekształcenia typowe dla wizji immersyjnej, wprowadzono metrykę IV-PSNR [4]. Metryka IV-PSNR jest koncepcyjnie oparta na PSNR (*Peak signal-to-noise ratio*) ale zawiera znaczne modyfikacje pozwalające na uwzględnienie zniekształceń występujących w obrazie powstałym w wyniku przerzutowania danych z kamer źródłowych w pozycję kamery wirtualnej. Najistotniejsze modyfikacje to dodanie narzędzia uwzględniającego globalną różnicę między komponentami (GCR – Global Component Difference) oraz mechanizmu kompensującego przesunięcie odpowiadających sobie pikseli (CPS – Corresponding Pixel Shift).

Jak przedstawiono na rys. 2, podstawowa wersja metryki IV-PSNR wykazuje wyższą korelację z oceną jakości subiektywnej niż inne metryki. Jednakże nie wykorzystuje ona analizy wizji w dziedzinie czasu. W niniejszym artykule zaproponowaliśmy uwzględnienie w obliczeniach metryki IV-PSNR analizy podobieństwa między charakterystykami czasowymi dwóch porównywanych sekwencji wizyjnych.

2. ZAPROPONOWANA METRYKA JAKOŚCI

Zaproponowane podejście opiera się na rozszerzeniu metryki IV-PSNR o informacje dotyczące zmian w charakterystyce czasowej mierzonej sekwencji. Najbardziej powszechne metryki jakości stosowane w aplikacjach wizji wszechogarniającej (PSNR, SSIM, IV-PSNR oraz WS-PSNR [13]) działają na pojedynczym obrazie, a jakość sekwencji obliczana jest przez proste uśrednienie jakości obliczonej dla poszczególnych klatek sekwencji wizyjnej. Niestety, takie podejście sprawia, że wspomniane metryki są niewrażliwe na pogorszenie spójności ruchu. Niektóre artefakty związane z czasem, takie jak migotanie krawędzi obiektów, są bardzo irytujące dla widza i powinny być uwzględnione w mierzonej jakości obiektywnej.

Zaproponowane podejście obejmuje wykorzystanie analizy pola ruchu [2] do oceny spójności ruchu sekwencji. Pole ruchu jest określane za pomocą gęstego przepływu optycznego (ang. DOF, *dense optical flow*) z wykorzystaniem opisanego w [6] algorytmu. Obliczenia przepływu optycznego są wykonywane dla obu sekwencji (referencyjnej i ocenianej), a ich wynik (w postaci wektorów ruchu) traktowany jest jako czwarty składnik obrazu. Wizualizacja różnicy pomiędzy polami ruchu obliczonymi dla obrazu oryginalnego i zniekształconego jest przedstawiona na rysunku 3.

Wszystkie rozważane metryki opierają się na PSNR, dlatego najważniejszym etapem wyznaczania każdej z nich jest obliczenie sumy kwadratów różnic (ang. SSD, *sum of squared differences*).

$$PSNR_c = 10 \log_{10} \frac{W \cdot H \cdot (2^b - 1)^2}{SSD_c}, \quad (1)$$

gdzie: $PSNR_c$ – wartość miary PSNR dla składowej c ; W , H – wysokość i szerokość obrazu; b – głębokość bitowa; SSD_c – wartość SSD dla składowej c .

Obliczenie wspomnianych kwadratów różnic (SD) dla składników obrazu jest proste (ponieważ wartości składników koloru są skalarami) i dla pojedynczego punktu danej składowej barwnej wygląda następująco:

$$SD_c(y, x) = (I_c(y, x) - J_c(y, x))^2, \quad (2)$$

gdzie: $SD_c(y, x)$ – wartość SD dla punktu o współrzędnych (y, x) należącego do składowej c ; $I_c(y, x)$ – próbka o współrzędnych (y, x) należącego do składowej c obrazu referencyjnego (I); $J_c(y, x)$ – próbka o współrzędnych (y, x) należącego do składowej c obrazu ocenianego (J).

W przypadku miary PSNR suma kwadratów różnic (SSD_c) jest obliczana przez wyznaczenie wartości $SD_c(y, x)$ dla całej składowej barwnej a następnie ich zsumowanie.

Natomiast dla wektorów ruchu SD jest obliczane jako kwadrat długości wektora powstałego przez obliczenie różnicy pomiędzy wektorem ruchu w obrazie referencyjnym (\vec{V}_I) a wektorem ruchu w obrazie ocenianym (\vec{V}_J):

$$SD_M(y, x) = \left| \vec{V}_I(y, x) - \vec{V}_J(y, x) \right|^2. \quad (3)$$

Obrazy wejściowe mogą cechować się różnymi wartościami głębi bitowej. Obecnie najczęściej spotykanymi wartościami są 8 lub 10 bitów na próbkę. W celu zachowania spójności wyznaczanych metryk i niezależnienia ich od głębi bitowej zastosowano korekcję wartości $SD_M(y, x)$ zależną od głębi bitowej. Jako wartość bazową przyjęto 10 bitów na próbkę, zaś sama operacja korekty została wyrażona w następujący sposób:

$$SD_M^{\wedge}(y, x) = SD_M(y, x) \cdot 2^{(b-10)}. \quad (4)$$

gdzie: $SD_M^{\wedge}(y, x)$ – skorygowana wartość SD dla pola ruchu; b – głębia bitowa.

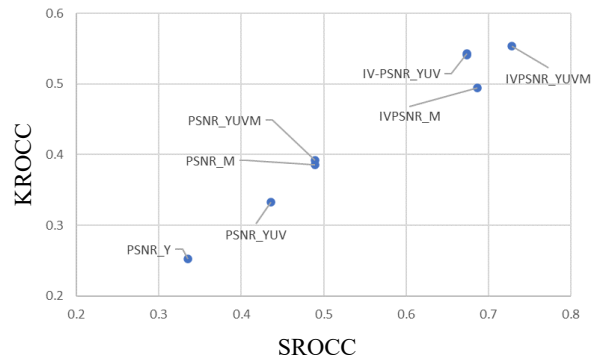
3. EKSPERYMENT

Zaproponowane podejście zostało ocenione na podstawie wyników „MPEG Call for Proposals on 3DoF+ Visual” [8], zawierających sekwencje wielowidokowe zakodowane przy użyciu siedmiu różnych technik kodowania wizji wszechogarniającej. Szczegółowa metodologia eksperymentu jest opisana w rozdziale VII artykułu [4], ponieważ została wcześniej wykorzystana do oceny podstawowej metryki IV-PSNR. Ze względu na dużą złożoność obliczeniową związaną z wyznaczaniem gęstego przepływu optycznego oraz dostępność sekwencji testowych, ocena została przeprowadzona na jednej grupie obrazów (ang. GOP, *group of pictures*). Propozycja została porównana z niezmiennym IV-PSNR, ponieważ zapewnia on najwyższą korelację z jakością subiektywną w aplikacjach wizji wszechogarniającej (rys. 2). Ponadto zaprezentowaliśmy wyniki dla zmodyfikowanego i niezmodyfikowanego PSNR, pokazując wszechstronność proponowanej modyfikacji (wykorzystanie informacji o ruchu). W eksperymencie metryki porównywano za pomocą dwóch powszechnie znanych i stosowanych współczynników korelacji: SROCC (Spearman rank correlation coefficient) i KROCC (Kendall rank correlation coefficient) [10]. Łącznie porównano siedem wariantów metryk:

- trzy metryki znane z literatury: **PSNR_Y** (PSNR wyłącznie dla składowej luminancji), **PSNR_{YUV}** (średnia ważona PSNR dla trzech składowych barwnych) i **IV-PSNR_{YUV}** (podstawowa wersja IV-PSNR, opisana w [4], uwzględniająca trzy składowe barwne),
- PSNR i IV-PSNR obliczone na sekwencjach pola ruchu, nieuwzględniające tekstury (ostatni rząd rys. 3): **PSNR_M** i **IV-PSNR_M**,
- Czterokomponentowe warianty miar PSNR i IV-PSNR, gdzie wektory ruchu traktowane są jako czwarty komponent (M), wraz z luminancją (Y) i obiema chrominancjami (Cb, Cr): **PSNR_{YUVM}** i **IV-PSNR_{YUVM}**.

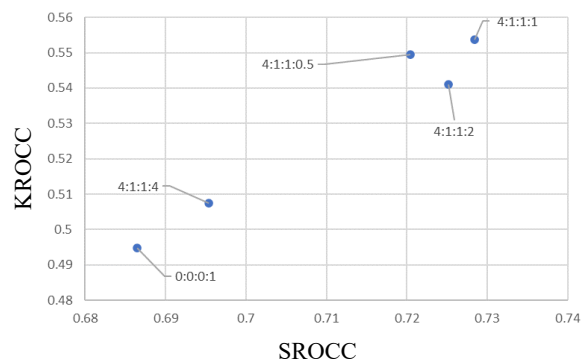
Wartość PSNR_{YUV} wyznaczana była poprzez uśrednienie wartości PSNR dla luminancji i obu chrominancji, z wagami – odpowiednio – 6:1:1 (zgodnie z [7]). W przypadku IV-PSNR_{YUV} wagi te zostały ustalone na 4:1:1 (zgodnie z [4]).

Korelacja między jakością subiektywną (obliczaną jako MOS, ang. *mean opinion score*) a wszystkimi testowanymi wariantami metryk jest przedstawiona na rys. 4. Jak pokazano, czterokładnikowy IV-PSNR_{YUVM} osiągnął najwyższą korelację z jakością subiektywną, zarówno pod względem SROCC, jak i KROCC. Użycie tylko wektorów ruchu (IV-PSNR_M) jest mniej skuteczne i charakteryzuje się podobną efektywnością do IV-PSNR obliczonego dla trzech składowych koloru (IV-PSNR_{YUV}).



Rys. 4. Korelacja pomiędzy subiektywną oceną jakości a jakością mierzona obiektywnie; porównanie testowanych wariantów metryk PSNR i IV-PSNR.

Jak przedstawiono, dodanie wektorów ruchu jako czwartej składowej jest efektywne również dla metryki PSNR. W przypadku PSNR podobna korelacja z MOS może być osiągnięta poprzez użycie tylko wektorów ruchu. Przyczyną takiego zjawiska jest obliczenie samego pola ruchu – ruch jest szacowany poprzez wyszukiwanie podobnych bloków o rozmiarze 11×11 w kolejnych klatkach sekwencji, a dla wizji wszechogarniającej analiza oparta na blokach jest bardziej efektywna niż oparta na pikselach (por. PSNR vs. IV-PSNR).



Rys. 5. Korelacja pomiędzy subiektywną oceną jakości a jakością mierzona obiektywnie; porównanie pięciu wariantów metryki IV-PSNR_{YUVM}.

Na rys. 5 przedstawiono porównanie pięciu wariantów IV-PSNR_{YUVM}. Warianty różnią się wagami przyporządkowanymi do poszczególnych składowych (Y:Cb:Cr:M). Cztery z nich (4:1:1:4, 4:1:1:2, 4:1:1:1 i 4:1:1:0.5) używają domyślnych wag dla luminancji i obu

chrominancji i różnią się w zestawie wag dla wektorów ruchu. Wariant 0:0:0:1 używa tylko wektorów ruchu, co odpowiada wartości IV-PSNR_M. Jak przedstawiono, najbardziej efektywny schemat ważenia równoważy wagi między wektorami ruchu a oboma chrominancjami. Podsumowując wyniki przedstawionych eksperymentów, stwierdzić można, iż wagi przyporządkowane do poszczególnych składowych Y:Cb:Cr:M równe 4:1:1:1 dają najwyższą korelację z ocenami subiektywnymi.

4. PODSUMOWANIE

W artykule przedstawiono efektywną modyfikację metryki IV-PSNR [4], która obejmuje analizę porównywanych sekwencji w dziedzinie czasu poprzez pomiar spójności ruchu, tj. podobieństwo gęstego pola ruchu w sekwencji referencyjnej i testowanej.

Propozycja opiera się na dodaniu wektorów ruchu jako czwartej składowej, uzupełniając trzy składowe barwne: Y, C_B i C_R. Wyniki eksperymentów pokazały, że proponowana modyfikacja IV-PSNR pozwala na zwiększenie korelacji między jakością obiektywną a subiektywną w ocenie jakości wizji wszechogarniającej. Biorąc pod uwagę, że propozycja poprawia metrykę IV-PSNR, a podstawowa metryka IV-PSNR przewyższa inne najnowsze metryki oceny jakości obiektywnej, można stwierdzić, że zaproponowana metryka jest wartościowa i bardzo praktyczna do zastosowania w rzeczywistych systemach wizji wszechogarniającej.

PODZIĘKOWANIA

Praca finansowana ze środków przyznanych przez Ministerstwo Nauki i Szkolnictwa Wyższego.

LITERATURA

- [1] Boyce Jill, Dore Renaud, Dziembowski Adrian, Fleureau Julien, Jung Joel, Kroon Bart, Salahieh Basel, Kumar Malamal Vadakital Vinod, Yu Lu. 2021. "MPEG Immersive Video coding standard". *Proceedings of the IEEE*, 119 (9): 1521-1536.
- [2] Brox Thomas, Malik Jitendra. 2011. Large Displacement Optical Flow: "Descriptor Matching in Variational Motion Estimation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (3): 500-513.
- [3] Domański Marek, Stankiewicz Olgierd, Wegner Krzysztof, Grajek Tomasz. 2017. "Immersive visual media — MPEG-I: 360 video, virtual navigation and beyond". *24th International Conference on Systems, Signals and Image Processing*, Poznań, Polska.
- [4] Dziembowski Adrian, Mieloch Dawid, Stankowski Jakub, Grzelka Adam. 2022. "IV-PSNR – The objective quality metric for immersive video applications". *IEEE Transactions on Circuits and Systems for Video Technology*, 32 (11): 7575-7591.
- [5] Dziembowski Adrian, Mieloch Dawid. 2022. "Nowe techniki kompresji wizji dla rzeczywistości wirtualnej – MPEG Immersive Video". *Przegląd Telekomunikacyjny - Wiadomości Telekomunikacyjne*. 4: 116-123.
- [6] Farneböck Gunnar. 2003. "Two-frame motion estimation based on polynomial expansion". *Scandinavian conference on Image analysis*, Berlin, Niemcy.
- [7] Huang Yue, Qi Honggang, Li Bin, Xu Jizheng. 2014. "Adaptive weighted distortion optimization for video coding in RGB color space". *IEEE International Conference on Image Processing, ICIP*, Paryż, Francja.
- [8] ISO/IEC. 2019. "Call for Proposals on 3DoF+ Visual". *ISO/IEC JTC1/SC29/WG11 MPEG N18145*, Marakesz, Maroko.
- [9] Liu Tsu-Ming, Ju Chi-Cheng, Huang Yu-Hao, Chang Tsui-Shan, Yang Kai-Min, Lin Yi-Ting. 2016. "A 360-degree 4K×2K panoramic video recording over smart-phones". *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Seattle, USA.
- [10] Myers Jerome, Well Arnold. 2003. "Research design and statistical analysis". *Lawrence Erlbaum Associates*, Londyn, UK.
- [11] The Netflix Tech Blog. 2016. Dostęp online: <https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652>
- [12] Stankiewicz Olgierd, Domański Marek, Dziembowski Adrian, Grzelka Adam, Mieloch Dawid, Samelak Jarosław. 2018. "A free-viewpoint television system for horizontal virtual navigation". *IEEE Transactions on Multimedia*, 20 (8): 2182-2195.
- [13] Sun Yule, Lu Ang, Yu Lu. 2017. "Weighted-to-Spherically-Uniform Quality Evaluation for Omnidirectional Video". *IEEE Signal Processing Letters*, 24 (9): 1408-1412.
- [14] Wang Zhou, Bovik Alan, Sheikh Hamid, Simoncelli Eero. 2004. "Image quality assessment: From error measurement to structural similarity". *IEEE Transactions on Image Processing*, 13 (4): 600-612.
- [15] Wien Mathias, Boyce Jill, Stockhammer Thomas, Peng Wen-Hsiao. 2019. "Standardization status of immersive video coding". *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9 (9): 5-17.