



Wrocław, 10-12 czerwca 2013

OBIEKTYWNA, STRUMIENIOWA METRYKA JAKOŚCI VIDEO 3D-HD

Streszczenie: Wzrost zainteresowania usługami wideo w technologii 3D powoduje również zwiększenie zapotrzebowania na prognozowanie jakości postrzeganej przez użytkownika. Jest ono niezbędne dla optymalizacji parametrów sieciowych. Nie jest możliwe proste przełożenie metod i metryk stosowanych dotychczas dla wideo 2D ze względu na nowe czynniki, związane przede wszystkim z efektem 3D. W artykule zaprezentowano metrykę (typu strumieniowego) jakości QoE dla wideo 3D-HD zniekształconego przez kompresję i straty pakietów.

1. WSTĘP

Warunkiem korzystania przez użytkowników z usług zaproponowanych przez operatora telekomunikacyjnego jest ich satysfakcjonująca jakość (ang. *Quality of Experience*, QoE). Metodologia pomiaru tej jakości dla usług multimedialnych jest od lat przedmiotem badań prowadzonych przez zespoły naukowców z całego świata. Najczęściej wynikiem końcowym tych prac są metryki jakości, których celem jest jak najlepsza prognoza poziomu satysfakcji użytkownika z korzystania z usługi multimedialnej. Przy czym przez metrykę rozumiemy funkcję, która określa wartość MOS (Mean Opinion Score) jakości QoE w zależności od wybranych parametrów łańcucha dostarczania treści. Informacja dostarczona przy użyciu metryki ma wymierny wpływ na planowanie i optymalizację przedsięwzięcia biznesowego, a także dla jego późniejszego sukcesu komercyjnego.

Coraz większym zainteresowaniem aktorów rynku cieszą się usługi multimedialne wideo 3D, czyli takie które umożliwiają percepcję głębi niemożliwą do osiągnięcia za pomocą metod wykorzystywanych w technologii 2D takich jak okluzja, cieniowanie, perspektywa geometryczna. Potencjalnymi obszarami zastosowań obrazów trójwymiarowych są: edukacja, rozrywka, medycyna chirurgiczna, videokonferencje i wiele innych. W najbliższych latach znacznie wzrośnie zainteresowanie treściami tego typu. Taki wniosek można wysnuć z prognoz przedstawionych w Cisco Visual Networking Index: Forecast and Methodology [1]. Przewiduje się, że latach 2011-2016 średnia roczna stopa wzrostu dla usługi wideo 3D na życzenie (ang. *Video-on-Demand 3D*, *VoD 3D*) będzie równa 109%. Oznacza to, że globalny ruch IP generowany przez transmisję VoD 3D wzrośnie z 2 PB na miesiąc w roku 2011 do 82 PB na miesiąc w 2016 roku. Podobny trend utrzymuje się w raporcie organizacji NPD DisplaySearch - 3D Display Technology and Market Forecast Report [2]. Przewiduje się, że

sprzedaż urządzeń wyświetlających 3D wzrośnie z 700 tysięcy sztuk w 2009 roku do 196 milionów w 2018 (średnia roczna stopa wzrostu to 75%). Te dane w zupełności uzasadniają przyczyny zainteresowania operatorów odpowiednimi metrykami jakości QoE dla wideo 3D. Należy w tym miejscu wspomnieć, że choć występuje duże podobieństwo łańcuchów dostarczania treści 2D i 3D to jednak ze względu na specyficzną naturę treści trójwymiarowych, ich zakłócenia i zniekształcenia występujące w kolejnych etapach łańcucha dostarczania są diametralnie inne, od tych znanych dla łańcucha 2D [3]. Z tej przyczyny zastosowanie metryk zaprojektowanych dla wideo 2D w przypadku 3D będzie co najmniej nieefektywne jeśli w ogóle możliwe.

W artykule zaprezentowano metrykę typu strumieniowego (ang. *bitstream metric*) służącą do pomiaru jakości wideo 3D w czasie rzeczywistym bez konieczności pełnego dekodowania. Uwzględnia ona wpływ kompresji i strat pakietów na jakość stereoskopowego wideo FullHD prezentowanego na migawkowym wyświetlaczu 3D. Do budowy metryki wykorzystano wyniki badań i wnioski uzyskane w wyniku przeprowadzenia testu jakości subiektywnej QoE.

Struktura artykułu jest następująca: w sekcji 2 przedstawiono aktualny stan wiedzy tj. metryki zaproponowane przez zespoły naukowe prowadzące badania w innych ośrodkach. W sekcji 3 omówiono pokrótce najważniejsze elementy przeprowadzonego eksperymentu – w tym sposób uzyskania materiału testowego. Opis metryki, jej składników i sposób wyznaczania jakości QoE w skali MOS zaprezentowano w sekcji 4. Podsumowanie zawarto w sekcji 5.

2. STAN WIEDZY

Tematyka badań Quality of Experience dla wideo 3D cieszy się coraz większym zainteresowaniem zespołów naukowych. Poniżej w skrócie zaprezentowano najważniejsze wyniki tych badań.

Xing i in. [4] zaproponowali metrykę, której współczynnik korelacji Pearsona jest równy 0,88. Jakość jest tutaj rozumiana w aspekcie percepcji poziomu przeników przez użytkownika. Przez przenik rozumiemy niedoskonałość separacji widoków powodującą, że obraz przeznaczony dla jednego oka jest również widoczny dla drugiego oka. Do oszacowania tego subiektywnego poziomu wykorzystano trzy główne czynniki: obiektywny poziom przeników, rozstaw obiektywów oraz charak-

terystykę treści sekwencji. Należy zaznaczyć, że w prezentowanych badaniach korzystano z reprezentacji trójwymiarowej w postaci kolor plus mapa głębi. W teście subiektywnym wzięło udział 28 osób.

Zhu i Wang [5] wykorzystali najnowsze osiągnięcia dotyczące układu wzrokowego człowieka, ściśle związane ze sferą fizjologiczną i psychologiczną. Przeanalizowali oni kilka głównych zależności związanych z tym układem, a dotyczących percepcji wideo 3D. Wśród nich funkcję czułości kontrastowej (ang. *Contrast Sensitivity Function*), funkcję percepcji głębi (ang. *Depth Perception*), maskowanie (ang. *Masking*) oraz mechanizm wielokanałowy (ang. *Multi-channel Mechanism*). Na tej podstawie zaproponowali wielokanałowy model wizyjny oparty o dekompozycję falkową dla sekwencji 3D. W ten sposób uzyskana przez nich metryka uzyskała korelację R^2 na poziomie ok. 0,65. Nie podano liczby testerów biorących udział w ewaluacji.

Joveluro i in. [6] do oceny jakości wideo 3D zastosowali metrykę Perceptual Quality Metric (PQM) dla wideo 2D, która według autorów jest bardziej efektywna niż Video Quality Metric (VQM). Określili ją jako bardziej czułą na niewielkie zmiany w postaci zakłóceń obrazu, ponieważ obliczanie błędów dokonywane jest już od poziomu pikseli. Weryfikacja została dokonana przy pomocy serii badań subiektywnych. Metryka PQM oblicza zakłócenia luminancji i kontrastu. Do badań, w których uczestniczyło 30 testerów, użyto czterech sekwencji o rozdzielczości 960x540 w reprezentacji kolor plus mapa głębi. Korelacje Pearsona uzyskane dla kolejnych sekwencji były równe około 0,97.

Obecny stan wiedzy w przedmiotowym zakresie nie zawiera badań zorientowanych na budowę metryk strumieniowych, możliwych do zastosowania w czasie rzeczywistym, bez konieczności pełnego dekodowania. Tego typu rozwiązania są najbardziej oczekiwane przez operatorów. Ponadto prowadzone obecnie badania skoncentrowane są wokół reprezentacji kolor plus mapa głębi, a nie uwzględniają reprezentacji stereoskopowej typu lewy widok plus prawy widok (ang. *side by side*).

3. OPIS EKSPERYMENTU

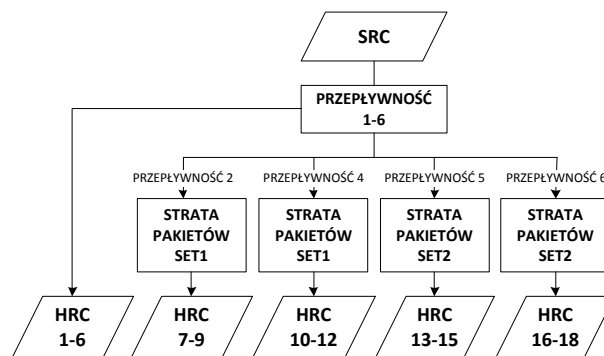
W tej sekcji przedstawiono skrócony opis eksperymentu, którego wyniki posłużyły do zaprojektowania przedmiotowej metryki jakości. Szczegółowy opis sposobu przygotowania sekwencji testowych, które podlegały ocenie przez testerów znajduje się w [4].

3.1. Sekwencje źródłowe

Sekwencje źródłowe to krótkie filmy trójwymiarowe w rozdzielczości 4K udostępnione w ramach współpracy w Video Quality Experts Group (VQEG) przez NTT z Japonii. Podzielono je na sceny, z których wybrano 9 najbardziej różnorodnych a następnie dokonano konwersji do rozdzielczości FullHD w trybie „*up side down side*” (obraz dla lewego oka i obraz dla prawego oka znajdują się jednej ramce jeden nad drugim). W ten sposób uzyskano sekwencje w rozdzielczości 1920 x 2160 pikseli.

3.2. Sekwencje testowe (PVS)

Sekwencje testowe zostały wygenerowane przez zmodyfikowanie sekwencji źródłowych przy użyciu 18 różnych układów referencyjnych (HRC, *Hypothetical Reference Circuits*) przedstawionych na rysunku 1.



Rys. 1 Generacja HRC

Wszystkie sekwencje źródłowe zostały zakodowane dla 6 różnych przepływności – HRC 1-6 (1 Mbit/s, 2 Mbit/s, 3 Mbit/s, 4 Mbit/s, 5 Mbit/s oraz 50 Mbit/s). Dodatkowo utworzono trzy poziomy strat pakietów (od 0 do 10^{-2}), które zostały dodane do czterech z sześciu poziomów przepływności sekwencji (HRC 7-18). W ten sposób powstały 162 sekwencje użyte następnie w trakcie testów (PVS). Przykładowe ramki sekwencji źródłowe przedstawiono na rysunku 2.



Rys. 2 Przykładowe ramki sekwencji użytych w eksperymencie (pierwszy wiersz – „*digest*”, drugi – „*okuagi*”, trzeci – „*okunai*”)

3.1. Przebieg eksperymentu

W badaniu wzięło udział łącznie 30 osób, które nie zajmują się w sposób profesjonalny usługami multimedialnymi. Kolejnym warunkiem udziału w eksperymencie były pozytywny wynik percepcji głębi wykonanego przy pomocy RANDOT Test. Wykonano także test ostrości wzroku oraz test na występowanie daltonizmu. Każdy tester wykonywał badanie indywidualnie według scenariusza metody ACR (*Absolute Category Rating*) opisanej w Rekomendacji ITU-T P.910 [8]. Po obejrzeniu trwającej 10 s sekwencji wideo tester miał 14 s na wskazanie subiektywnej oceny jakości sekwencji w skali MOS; po tym czasie na wyświetlaczu prezentowano kolejną sekwencję do oceny.

Środowisko testowe przygotowano zgodnie z Rekomendacjami ITU-R BT.500-12 [9], ITU-T P.910 [8] i ITU-R BT.1438 [10]. W teście użyto wyświetlacza 3D BENQ 24 wraz z okularami migawkowymi. Na rysun-

ku 3 zaprezentowano pomieszczenie, w którym przeprowadzono test.

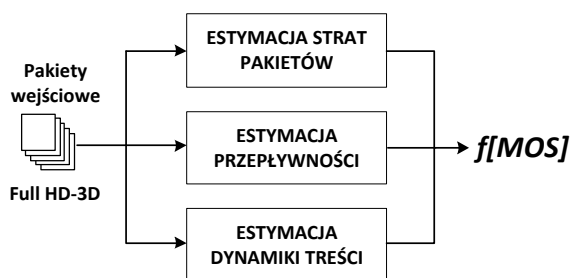


Rys. 3 Środowisko testowe w Katedrze Telekomunikacji AGH

4. OPIS METRYKI

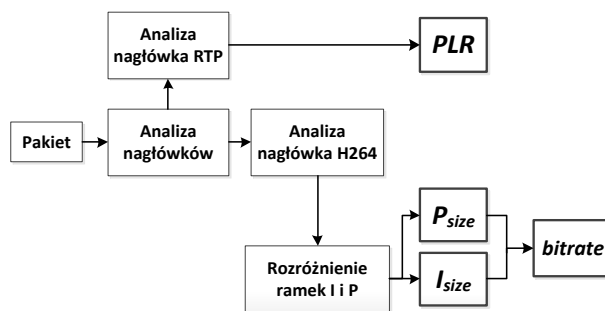
Do budowy metryki zostały wykorzystane parametry ekstrahowane ze strumienia danych z pomocą odpowiedniego oprogramowania. Parametry te, po związaniu ich modelem liniowym, wyznaczają jakość sekwencji 3D.

Na rysunku 4 przedstawiono schemat ideowy budowy metryki. Moduł do pomiaru jakości postrzeganej przez użytkownika QoE analizuje wszystkie pakiety, które otrzymuje klient. Następnie wykrywa i estymuje straty pakietów, estymuje przepływność i dynamikę treści. Jak wykazano w [4], ocena jakości QoE dla 3D jest szczególnie wrażliwa na rodzaj przesyłanej treści. Ze względu na zastosowanie kodeka opartego na predykcji ruchu, estymacja dynamiki treści może być stosunkowo łatwo wykonana za pomocą porównania rozmiarów ramek P i I.



Rys. 4 Składowe metryki QoE dla video 3D

Na rysunku 5 przedstawiono sposób ekstrakcji składników metryki. Na podstawie dostarczonej sekwencji pakietów dokonywana jest analiza nagłówków. Analiza nagłówka RTP dostarcza informacji o straconych pakietach, która jest wykorzystywana do obliczenia *PLR*. Z kolei analiza nagłówka strumienia H264 pozwala na rozróżnienie ramek I oraz P, a następnie wyznaczenie ich rozmiarów. Na podstawie tych informacji możliwe jest wyznaczenie składnika *bitrate*.



Rys. 5 Budowa modułu do pomiaru jakości QoE dla 3D za pomocą zaprezentowanej metryki

Ze względu na duży zakres przepływności, metryka została wyznaczona odrębnie dla dwóch grup: przepływności mniejszych od 4 Mbit/s (1) oraz przepływności większych od 4 Mbit/s (2).

$$f_{low}[MOS] = a_{low}PLR + b_{low}bitrate + c_{low} \quad (1)$$

$$f_{high}[MOS] = a_{high}PLR + b_{high}Psize + c_{high} \quad (2)$$

gdzie:

$$PLR = \frac{(\text{liczba straconych pakietów}) \cdot 100}{\text{liczba pakietów}}$$

$$bitrate = 8 \times (29 \times Psize + Isize)$$

Psize – średni rozmiar ramki P

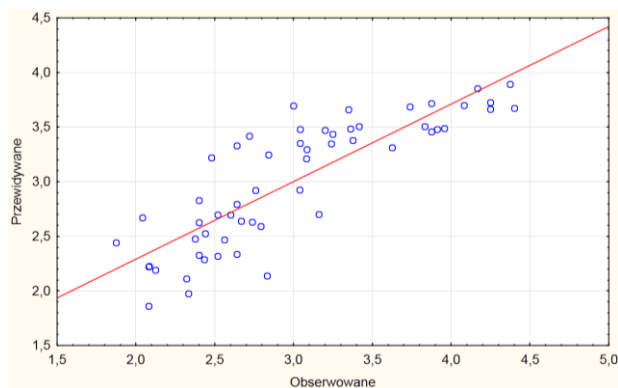
Isize – średni rozmiar ramki I

Współczynniki *a*, *b* i *c* zostały wyznaczone na podstawie wyników badań subiektywnych, które zostały opisane w sekcji 3.1. W tabeli 1 zamieszczono wartości tych współczynników dla obu grup przepływności.

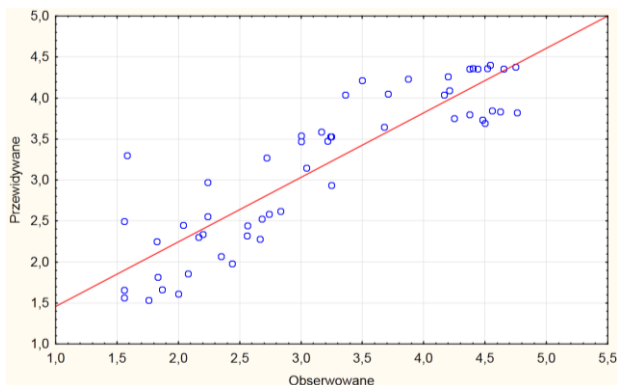
Tab. 1. Wyznaczone wartości współczynników *a*, *b* i *c*

	low	high
<i>a</i>	-0,51	-1,06
<i>b</i>	3,93	-0,28
<i>c</i>	5,58	3,11

Dalej, na rysunku 6 przedstawiono wykres zależności wartości MOS przewidywanych względem wartości obserwowanych wraz z modelem dla przepływności mniejszych niż 4 Mbit/s.



Rys. 6 Wykres wartości MOS przewidywanych względem obserwowanych dla przepływności poniżej 4 Mbit/s



Rys. 7 Wykres wartości MOS przewidywanych względem obserwowanych dla przepływności powyżej 4 Mbit/s

Z kolei na rysunku 7 przedstawiono taki wykres dla przepływności większych niż 4 Mbit/s. Uzyskana korelacja R^2 dla niższych przepływności jest równa 0,71, natomiast dla wyższych – 0,89. Ujednolicenie modelu dla całego spektrum przepływności obniżyłoby jego dokładność.

5. PODSUMOWANIE

Projektowanie metryk jakości postrzeganej przez użytkownika dla wideo 3D jest zadaniem bardzo trudnym ze względu na różnorodność czynników wpływających na percepcję efektu 3D (np. jakość obrazu, przeniki, zniekształcenia geometryczne, efekt teatru kukiełek, efekt teksturowy, efekt skrócenia perspektywy, a także przemęczenie wzroku itp.). W tym artykule zaprezentowano metrykę zbudowaną na podstawie parametrów ekstrahowanych ze strumienia danych. Tym samym przedstawione rozwiązanie należy do grupy metryk strumieniowych, bez referencji, które w obliczu przedstawionego stanu wiedzy jest konkurencyjne i oryginalne. Oznacza to, że w istniejącym stanie wiedzy do tej pory nie zbudowano metryki jakości 3D dla wideo o rozdzielczości FullHD w reprezentacji stereoskopowej dla lewego i prawego widoku – obecnie najbardziej popularnej. Z praktycznego punktu widzenia niewątpliwą zaletą jest możliwość użycia metryki w czasie rzeczywistym, bez konieczności pełnego dekodowania treści.

Dotychczasowe wyniki są bardzo obiecujące, dlatego konieczne jest przeprowadzenie dalszych badań subiektywnych, które mogłyby zwiększyć wiedzę na temat zależności jakości QoE od rodzaju prezentowanych treści. Podobnie, bardziej pogłębiona analiza strumienia danych (np. wyznaczenie średnich parametrów kwantyzacji, statystyk dotyczących makrobloków, wpływu strat na ramki I, P, B), pozwoli na zwiększenie dokładności kolejnych modeli.

6. PODZIĘKOWANIA

Artykuł powstał w ramach Programu Operacyjnego Innowacyjna Gospodarka - projekt nr POIG.01.01.01-00-045/09-00 „Inżynieria Internetu Przyszłości” sfinansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego z Europejskiego Funduszu Rozwoju Regionalnego.

SPIS LITERATURY

- [1] Cisco Visual Networking Index: Forecast and Methodology, 2011-2016, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf
- [2] NPD DisplaySearch, 3D Display Revenues Forecast to Reach \$22B by 2018; 3D-Ready TV Shipments to Reach 64M Units, http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/100104_3d_display_revenues_forecast_to_reach_22_billion_by_2018.asp
- [3] D. Juszka, „Badania subiektywne jakości QoE dla usług telewizji trójwymiarowej”, *Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne*; ISSN 1230-3496, 2011, R. 84 nr 6 s. 633–636
- [4] L. Xing, J. You, T. Ebrahimi, A. Perkiş, “A perceptual quality metric for stereoscopic crosstalk perception”, 17th IEEE International Conference on Image Processing (ICIP), str. 4033-4036, 2010
- [5] Z. Zhu, Y. Wang, “Perceptual distortion metric for stereo video quality evaluation”, *WSEAS Trans. Sig. Proc.* 5, 7, str. 241-250, 2009
- [6] P. Joveluro, H. Malekmohamadi, W. A C Fernando, A.M. Kondo, “Perceptual Video Quality Metric for 3D video quality assessment,” 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), str.1,4, 2010
- [7] D. Juszka, L. Janowski, Z. Papir, „Ocena jakości wideo 3D-HD zniekształconego przez kompresję i straty pakietów”, *Przegląd Telekomunikacyjny, Wiadomości Telekomunikacyjne*, ISSN 1230-3496 201,2 R. 85 nr 4 s. 294–297
- [8] ITU-T P.910, Subjective video quality assessment methods for multimedia applications, International Telecommunication Union Recommendation, 1999
- [9] ITU-R BT.500-12, Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union Recommendation, Rev. 12, 2009
- [10] ITU-R BT.1438, Subjective assessment of stereoscopic television pictures, International Telecommunication Union Recommendation, 2000