



Kraków, 27-29 czerwca 2016

BADANIE KORELACJI POMIĘDZY TECHNICZNYMI PARAMETRAMI OBRAZU WIDEO STEREOSKOPOWEGO (3D) A SUBIEKTYWNĄ OCENĄ TREŚCI WIDEO

A STUDY ON CORELATIONS BETWEEN TECHNICAL PARAMETERS OF STEREOSCOPIC 3D VIDEO IMAGE AND SUBJECTIVE ASSESSMENT OF VIDEO CONTENT

Streszczenie: Rozwój usług telekomunikacyjnych zmienia sposób korzystania z treści wideo. Duża dostępność różnorodnej informacji powoduje, że użytkownicy zaczynają dokonywać coraz bardziej precyzyjnych wyborów zgodnych ze swoimi preferencjami. Dążenie do personalizowania usług zaczyna być dostrzegane w obszarze badań jakości wideo. W artykule wykazano brak korelacji pomiędzy zainteresowaniem treścią, subiektywną oceną atrakcyjności wizualnej próbki i wrażeniem głębi a parametrami technicznymi: szczegółowością obrazu, jego dynamiką oraz gęstością cięć scen.

Abstract: The development of telecommunications services is changing the way we watch video content. The large availability of various information enable users to make more precise choices in line with their true preferences. This trend must be included in the area of Quality of Experience research. The paper shows that subjective assessment of personal interest in video content, its visual attractiveness and level of 3D effect impression is not correlated with objective parameters: spatial activity, temporal activity and mean scene cut density.

Słowa kluczowe: 3D, jakość, czynniki subiektywne, wideo

Keywords: 3D, quality, subjective factors, video

1. WSTĘP

Od ponad dwóch dekad intensywnie rozwijana jest metodyka pomiarów jakości multimedialnych postrzeganej przez użytkownika (ang. *Quality of Experience*, QoE). Mają one na celu poszukiwanie relacji pomiędzy technicznymi parametrami danej usługi a subiektywną oceną jakości wrażenia, jakie wywiera ona na użytkownika. Utylitarną motywacją dla tego rodzaju badań jest kontrola i kierunkowanie rozwoju metod kodowania i kompresji a także transmisji, aby jak najlepiej spełniały oczekiwania klientów dostarczając im pożądaną i optymalną jakość usługi. W wyniku takich badań powstają również metryki, które umożliwiają operatorom telekomunikacyjnym dostosowanie parametrów transmisji do wymagań użytkownika.

Jednym z obszarów usług, które stają się coraz bardziej popularne i zmieniają przyzwyczajenia użytkowników jest wideo na żądanie (ang. *Video-on-Demand*, VoD). Użytkownicy sami mogą świadomie wybierać

spośród rozmaitych, stale i szeroko dostępnych treści wideo, co więcej mogą je oglądać za pomocą różnorodnych urządzeń. Można zatem oczekiwać, że zmienia się zachowanie użytkowników – to indywidualny, wysoki poziom zainteresowania staje się główną motywacją do obejrzenia konkretnej treści, a nie jak poprzednio uwarunkowania zewnętrzne. Nie ulega wątpliwości, że takie zmiany muszą zostać uwzględnione w metodach badawczych jakości wideo postrzeganej przez użytkownika.

Obserwacja trendów występujących w serwisach VoD pozwala stwierdzić, że decyzja użytkownika o kontynuowaniu oglądania jest spowodowana nie tylko poziomem akceptacji jakości, ale także – a może przede wszystkim – wysokim poziomem zainteresowania treścią. Te obserwacje są zgodne z definicją QoE przyjętą w środowisku naukowym [1].

Przedstawiona ocena stanu rzeczy wskazuje na konieczność podjęcia badań uwzględniających czynnik zainteresowania treścią w selekcji materiału testowego do eksperymentów subiektywnych jakości wideo oraz w samej analizie wyników. Dodatkową motywacją jest fakt, że materiał wideo obecnie wykorzystywany w badaniach subiektywnych jest bardzo często wygenerowany czy nagrany specjalnie na potrzeby eksperymentów. Nie są to treści szczególnie interesujące dla uczestników badań i nie mogą być uznane za próbki treści, z którymi użytkownik rzeczywiście spotyka się korzystając na co dzień z usługi wideo-na-żądanie. Czas trwania takich materiałów to najczęściej nie więcej niż 15 s. Choć nie mają one ambicji posiadania fabuły czy prezentacji przebiegu jakiegoś wydarzenia, w takim odcinku czasu i tak nie jest możliwe oddanie rzeczywistego sposobu oglądania programu telewizyjnego. Co więcej – brak im cięć – zabiegu edycyjnego charakterystycznego dla profesjonalnych produkcji wideo. Inne elementy zwiększające artyzm produkcji to m.in. kadrowanie, oświetlenie, balans kolorów, ruch kamery czy efekty specjalne. Rozważając te różnice dla porządku jedynie należy wspomnieć o kwestiach obsady czy w ogóle gry aktorskiej.

W artykule zostały przedstawione wyniki badań subiektywnych zainteresowania treścią, atrakcyjnością wizualną próbki oraz wrażenia wywoływanego przez efekt 3D dla wideo zrealizowanego w technice stereoskopowej (osobny obraz dla lewego i prawego oka).

Zbadano korelacje pomiędzy tymi trzema cechami a rekomendowanymi przez ITU parametrami obiektywnymi selekcji materiału testowego do eksperymentów subiektywnych tj. szczegółowością obrazu wideo i jego dynamiką, a także parametrem własnym tj. gęstością cięć scen. Wystąpienie korelacji oznaczałoby, że możliwe jest zbudowanie modelu, który pozwoliłby na podstawie trzech parametrów technicznych (obiektywnych) określić czy użytkownik jest zainteresowany treścią wideo, lub czy jest ona dla niego atrakcyjna wizualnie lub czy sposób wykorzystania efektu 3D wywarł na nim wrażenie. Taki model pozwoliłby bez prowadzenia dodatkowych eksperymentów zwiększyć dokładność wyników subiektywnych badań jakości. Brak występowania korelacji uniemożliwia zbudowanie takiego modelu i dodatkowo uzasadnia konieczność opracowywania takich scenariuszy badań, które będą uwzględniały zjawisko wpływu treści wideo na subiektywną ocenę jakości. Ma to kluczowe znaczenie dla wiarygodności wyników badań w tym obszarze. Materiał testowy użyty w niniejszym badaniu pochodził z profesjonalnych, komercyjnych produkcji filmowych 3D, a czas trwania pojedynczej próbki był równy 30 ± 2 s.

Struktura artykułu jest następująca: w Sekcji 2 przedstawiono aktualny stan wiedzy, następnie w Sekcji 3 zaprezentowano opis przygotowania i przeprowadzenia eksperymentu. W Sekcji 4 znajduje się omówienie wyników badań.

2. STAN WIEDZY

P. Kortum i M. Sullivan przeprowadzili badania wpływu atrakcyjności (ang. *desirability*) treści na subiektywne oceny jakości użytkowników. W serii czterech badań wzięło udział 100 osób i oceniło 180 dwuminiutowych, dwuwymiarowych (2D) sekwencji zakodowanych na dziewięciu różnych poziomach jakości. Wykazano silną korelację pomiędzy atrakcyjnością treści i subiektywną oceną jakości [3]. Nie prowadzono podobnych badań dla wideo stereoskopowego (3D).

Badania nad uwzględnieniem wpływu estetyki i artystyzmu produkcji filmowych na oceny subiektywne uczestników eksperymentów są obecne w literaturze przedmiotu [7] **Błąd! Nie można odnaleźć źródła odwołania..** Zauważono, że odpowiednia edycja i praca kamery (cięcia, struktura narracji, kompozycja obrazu, żywe kolory) zwiększają zainteresowanie i wzmagają atrakcyjność, a to wpływa na ocenę jakości [5]. Ponadto zaobserwowano, że różnorodne pomniejsze defekty wizualne sekwencji mają duży kumulatywny wpływ na oceny jakości w skali MOS. Takim defektami są na przykład rozmazany ruch, szum analogowy, drgająca kamera, kiepska kompozycja, długie ujęcia (bez cięć) czy niezrozumiały sposób prezentacji. Te różnice mogą być zaobserwowane na przykład w bazie Consumer Digital Video Library (CDVL; www.cdvl.org), w której zostały udostępnione sekwencje (2D) wraz subiektywnymi ocenami jakości. Profesjonalnie zrealizowane nagranie wideo zawiera zdecydowanie więcej cięć niż amatorskie. Czas trwania ujęć i schemat ich występowania jest charakterystyczny dla gatunków filmów [2].

Dobór materiału testowego do eksperymentu subiektywnego wynika z celu badania. Wskazówki dotyczące eksperymentów w obszarze szeroko rozumianego wideo rozrywkowego [5] różnią się od tych dotyczących zorientowanych za zadanie [3]. Nawet w przypadku zastosowania zaprezentowanych kryteriów autorzy podkreślają konieczność instruowania testerów, by w swoich ocenach nie kierowali się treścią czy jakością gry aktorskiej. Zalecają oni wybór 8-10 klipów reprezentujących szerokie spektrum możliwych rodzajów treści. Należy unikać treści ofensywnych tj. scen kontrowersyjnych, nacechowanych przemocą, nieprzyzwoitością, polityką, religią, horrorami, zawierającymi fragmenty operacji medycznych czy spożywania i objawów po spożyciu substancji odurzających – więcej bardziej szczegółowych rekomendacji opracowano w [5]. M. Pinson i in. zwracają uwagę na kwestię występowania cięć, do tej pory często zaniedbywaną w środowisku badaczy jakości QoE. Obecnie Video Quality Experts Group rekomenduje aby do eksperymentów wybierać połowę klipów z cięciami scen, w tym jeden klip o dużej liczbie cięć (np. co 1-2 s) i połowę klipów bez cięć [5].

3. OPIS EKSPERYMENTU

3.1. Przygotowanie sekwencji testowych

Spośród dostępnych na rynku Bluray-3D wybrano dziewięć tytułów (*Niesamowity Spiderman*, *Dziki z natury*, *Piekielna zemsta*, *Hugo i jego wynalazek*, *Człowiek ze stali*, *Piraci z Karaibów: na nieznanym wodach*, *Polar Bear: Ice Bear*, *Wielki Gatsby*, *Hobbit: Niezwykła podróż*). Wśród nich znajdują się dwa filmy dokumentalne i siedem filmów fabularnych różnych gatunków, zrealizowanych przez różnych reżyserów. Łączny czas trwania wszystkich filmów to 19 godzin i 28 minut.

Nagrania pobrano z płyt Bluray-3D w najwyższej możliwej jakości, a następnie wycięto 96 sekwencji trwających 30 ± 2 s i dodatkowo 5 sekwencji, które wykorzystano w sesji treningowej. Każda z sekwencji charakteryzowała się ciągłością akcji – logicznie następującymi po sobie ujęciami. Następnie wszystkie sekwencje ułożono według losowej kolejności i podzielono na trzy zbiory po 32 sekwencje każdy. Z każdego takiego zbioru utworzono klip, czyli 32 sekwencje przeplatane trwającymi 15 s sekwencjami zawierającymi prośbę o udzielenie odpowiedzi na pytania, o których mowa w punkcie 3.4. Tę czynność powtórzono 4 razy, aby utworzyć 4 unikatowe zestawy losowo ułożonych sekwencji podzielone na 3 klipy każdy. Taki zabieg miał zapobiec tzw. efektowi kolejności wyświetlania. W przypadku filmów, których ramki były mniejsze niż 1920×1080 zastosowano kaszety (czarne pasy symetrycznie powyżej i poniżej ramki).

Dla każdej z sekwencji obliczono następujące parametry obiektywne:

- Aktywność przestrzenna (ang. *Spatial perceptual information*, SI) opisana w Rekomendacji ITU-T P.910 [9], która opisuje jak wiele detali zawiera najbardziej szczegółowa ramka w danej sekwencji

$$SI = \max_{time} \{std_{space} [Sobel(F_n)]\}$$

- Aktywność czasowa (ang. *Temporal perceptual information*, TI) opisana w Rekomendacji ITU-T P.910

[9], która opisuje największą ilość zmian w czasie jakie wystąpiły w danej sekwencji.

$$TI = \max_{time} \{std_{space}[F_n(i, j) - F_{n-1}(i, j)]\}$$

- Średnia wartość SI dla całej sekwencji jest średnią arytmetyczną z wartości SI dla kolejnych ramek

$$SI_{\text{średnie}} = \frac{\sum std_{space}[Sobel(F_n)]}{N}$$

- Średnia wartość TI dla całej sekwencji jest średnią arytmetyczną z wartości TI dla kolejnych ramek.

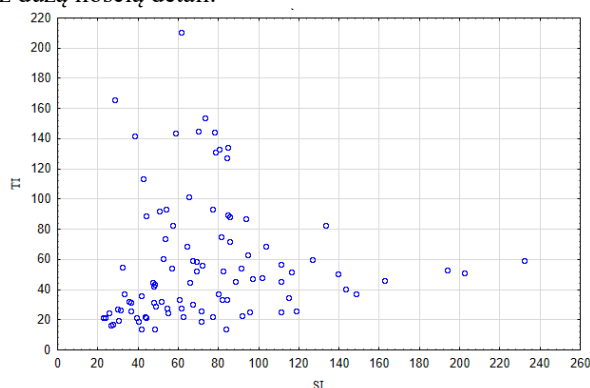
$$TI_{\text{średnie}} = \frac{\sum std_{space}[F_n(i, j) - F_{n-1}(i, j)]}{N}$$

- Gęstość cięć scen jest średnią arytmetyczną liczby ramek pomiędzy kolejnymi cięciami w danej sekwencji.

$$d = \frac{\sum(r_m - r_{m-1})}{M}$$

gdzie: F_n – płaszczyzna luminancji n -tej ramki; $std_{space}(\cdot)$ - odchylenie standardowe (\cdot); $Sobel(\cdot)$ – filtr Sobela; r_m – numer ramki w momencie m -tego cięcia sceny; N – liczba wszystkich ramek sekwencji; M – liczba wszystkich cięć w sekwencji

Obliczone wartości SI względem TI dla sekwencji użytych w eksperymencie przedstawiono na wykresie rozrzutu (Rys. 1). Przedstawia on, jak bardzo różnorodne pod względem parametrów obiektywnych były wybrane do badania sekwencje. Brak sekwencji w prawej górnej ćwiartce wykresu wynika z praktyki produkcji filmowej – ze względu na komfort widza i ograniczenia techniczne unika się bardzo dynamicznych scen z dużą ilością detali.



Rys. 1. Wykres rozrzutu wartości SI względem TI dla sekwencji użytych w eksperymencie

3.2. Środowisko testowe

Środowisko testowe zostało przygotowane zgodnie z Rekomendacjami ITU-R BT.500-12 [8], ITU-T P.910 [9]. Do prezentacji próbek użyto telewizora plazmowego 3D (PANASONIC TX-P42GT30). Zgodnie z rekomendacją ITU testerzy korzystali z okularów migawkowych 3D, a oglądali klipy siedząc w odległości równej trzem wysokościami ekranu. Odpowiedzi na pytania udzielali na papierowych arkuszach, aby uniknąć konieczności zdejmowania okularów migawkowych.

3.3. Charakterystyka testerów

W badaniu wzięło udział 28 osób, zatrudnionych z pomocą niezależnej agencji pośrednictwa pracy tymczasowej, za wynagrodzeniem, w żaden sposób niezwiązanych profesjonalnie z branżą usług multimedialnych, w szczególności z obszarem jakości wideo. Połowę osób uczestniczących w eksperymencie stanowiły kobiety.

Średni wiek osób był równy 23,1. Warunkiem wstępnym udziału był pozytywny wynik testu percepcji głębi (RANDOT Test) oraz testu weryfikującego widzenie barw (Ishihara Test).

3.4. Przebieg eksperymentu

Każdy tester wykonywał badanie pojedynczo, według scenariusza ACR (Absolute Category Rating) opisanego w Rekomendacji ITU-T P.910, przy czym w trakcie badania nie była oceniana jakość próbek a tylko ich subiektywne cechy (poziom zainteresowania wzbudzającego ich treścią, ich atrakcyjność wizualna oraz wrażenie wywoływane przez użyty w nich efekt 3D). Po obejrzeniu trwającej 30 ± 2 s sekwencji wideo tester miał 15 s na udzielenie odpowiedzi na trzy pytania:

1. Jak bardzo treść zaprezentowanego fragmentu filmu była dla Ciebie interesująca?
2. Jak bardzo zaprezentowany fragment filmu był dla Ciebie atrakcyjny wizualnie?
3. Jak oceniasz wpływ wrażenia głębi (3D) na swój odbiór zaprezentowanego fragmentu filmu?

W instrukcji do zadania każdy tester został poinformowany o interpretacji tych pytań. Na skali od 1 (w ogóle nieinteresująca) do 5 (bardzo interesująca) tester miał ocenić w tej kategorii przebieg akcji, sposób jego przedstawienia, fabułę. W odpowiedzi na drugie pytanie, na skali od 1 (w ogóle nieatrakcyjny) do 5 (bardzo atrakcyjny), tester miał ocenić jak bardzo aspekt estetyczny i artystyczny obejrzanego fragmentu filmu były dla niego atrakcyjne wizualnie. Proszony był o uwzględnienie w swojej ocenie barw, oświetlenia, tzw. pracy kamery, proporcji, a także szeroko rozumianej kreacji zaprezentowanego świata. Ostatnie pytanie, specyficzne dla techniki wyświetlania obrazu, miało na celu sprawdzenie percepcji głębi przez testera. Na skali od 1 (nie zauważyłem/am efektu 3D) do 5 (wywołał ogromne wrażenie) testerzy oceniali intensywność wrażenia głębi w obejrzanym fragmencie filmu.

Przed rozpoczęciem badania każdy tester miał możliwość zapoznania się z jego przebiegiem w trakcie sesji treningowej. Właściwe badanie odbywało się w trzech następujących po sobie sesjach. Po każdej sesji następowała krótka przerwa.

Na zakończenie badania każdy z uczestników samodzielnie uzupełniał ankietę osobową i podsumowującą udział w eksperymencie.

4. WYNIKI BADAŃ

Z analizy wyników wykluczono oceny jednej uczestniczki badania jako skrajnie odstające od całej grupy. Zestawienie obliczonych korelacji pomiędzy średnim poziomem zainteresowania treścią (INT), atrakcyjnością wizualną (ATR) i wrażeniem jakie wywołał efekt 3D (w3D) znajduje się w Tab. 1.

Tab. 1. Współczynniki korelacji pomiędzy parametrami subiektywnymi.

Korelacja	R ²
INT vs. ATR	0,55
INT vs. w3D	0,36
ATR vs. w3D	0,83

Pierwszą obserwacją jest duża korelacja pomiędzy ocenami atrakcyjności wizualnej i ocenami wrażenia głębi (wyniki statystycznie istotne $p < 0,05$). Może ona oznaczać, że testerzy mają trudności z odróżnieniem tych dwóch cech. Czynnikiem wpływającym na zdolność krytycznej oceny efektu 3D może być preferencja wyboru filmu zrealizowanego w tej technice. W ankiecie końcowej, na 27 uczestników tylko 11 osób mając do wyboru film zrealizowany w dwóch technikach, wybiera lub zdecydowanie wybiera film 3D. Niestety, liczba osób biorących udział w opisywanym eksperymencie jest zbyt mała by ocenić jak bardzo preferencja wyboru techniki realizacji filmu wpływa na zdolność do krytycznej oceny efektu 3D. Ta kwestia wymaga dodatkowych badań.

Zbadano korelacje pomiędzy średnimi ocenami subiektywnymi tzn. poziomem zainteresowania (INT), poziomem atrakcyjności wizualnej (ATR) i wrażeniem efektu 3D (w3D) oraz parametrami obiektywnymi czyli maksymalną wartością SI, średnią wartością SI, maksymalną wartością TI, średnią wartością TI oraz gęstością cięć scen. Tylko dla trzech par wyniki okazały się statystycznie istotne ($p > 0,05$). W pozostałych przypadkach uzyskano bardzo słabe korelacje, ale statystycznie istotne (wartości pogrubiono) co przedstawia Tab. 2. Oznacza to, że na podstawie obiektywnych parametrów nie można orzec czy dana sekwencja będzie dla uczestników interesująca czy atrakcyjna wizualnie. Podobnie zagęszczenie cięć wbrew intuicjom opisanym w aktualnym stanie wiedzy ma jedynie nieznaczny związek z badanymi cechami subiektywnymi.

Tab. 2. Współczynniki korelacji pomiędzy parametrami obiektywnymi a parametrami subiektywnymi.

Zmienna	SI	Średnie SI	TI	Średnie TI	Gęstość cięć
INT	-0,04	-0,16	0,28	0,19	-0,27
ATR	0,47	0,42	0,23	0,42	0,26
w3D	0,54	0,46	0,34	0,54	0,31

5. PODSUMOWANIE

W eksperymencie wykorzystano w formie sekwencji testowych fragmenty profesjonalnie zrealizowanych filmów fabularnych i dokumentalnych, co jest nowością w aktualnym stanie wiedzy. Ponadto użyte treści zostały nagrane w technice stereoskopowej (3D), a czas trwania każdej z użytych sekwencji był równy 30 ± 2 s.

Celem zaprezentowanego eksperymentu było zbadanie, czy istnieje zależność pomiędzy szczegółowością obrazu wideo, jego dynamiką i gęstością cięć scen a subiektywną oceną zainteresowania treścią, jak atrakcyjnością wizualną i wrażeniem jakie wywołuje efekt 3D. Nie potwierdzono związku pomiędzy gęstością cięć i badanymi cechami subiektywnymi wideo. Ponadto wskazano na niewielki związek tych parametrów z obiektywnymi cechami obrazu wykorzystywanymi powszechnie do selekcji materiału testowego tj. aktywnością przestrzenną (SI) oraz aktywnością czasową (TI).

Uzyskane wyniki oznaczają, że przy wyborze materiału testowego do eksperymentów badacze powinni kierować się nie tylko stosowanymi do tej pory parametrami obiektywnymi (SI oraz TI), ale także subiektywnymi ocenami treści wideo.

Przedstawione wyniki badań będą wykorzystywane do selekcji materiału w przygotowywanym eksperymencie pomiaru jakości wideo stereoskopowego.

6. PODZIĘKOWANIA

Autorzy dziękują Panu dr hab. M. Wierzchoniowi (Zakład Psychologii Eksperymentalnej Instytutu Psychologii Uniwersytetu Jagiellońskiego) za konsultacje i uwagi krytyczne do scenariusza eksperymentalnego.

Artykuł powstał w ramach projektu nr 15.11.230.195 sfinansowanego w konkursie „Grant Dziekański” organizowanym przez Wydział Informatyki, Elektroniki i Telekomunikacji Akademii Górniczo-Hutniczej im. St. Staszica w Krakowie.

LITERATURA

- [1] Möller, S., Raake, A. (Eds.). (2014). *Quality of Experience. Advanced Concepts, Applications and Methods*. T-Labs Series in Telecommunication Services, Springer. Retrieved from <http://www.springer.com/engineering/signals/book/978-3-319-02680-0>
- [2] Niu, Y., & Liu, F. (2012). What makes a professional video? a computational aesthetics approach. *IEEE Transactions on Circuits and Systems for Video Technology*, 22, 1037–1049. <http://doi.org/10.1109/TCSVT.2012.2189689>
- [3] Pinson, M. H., Boyd, K. S., Hooker, J., & Muntean, K. (2013). How to choose video sequences for video quality assessment?, *Proceedings of the Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-2013)*. Retrieved from <http://enpub.fulton.asu.edu/resp/vpqm/vpqm13/Papers/Paper4.pdf>
- [4] Fenimore, C., Libert, J., & Wolf, S. (1998). Perceptual effects of noise in digital video compression. In *140th SMPTE Technical Conference*. <http://doi.org/10.5594/M00301>
- [5] Pinson, M. H., Barkowsky, M., & Le Callet, P. (2013). Selecting scenes for 2D and 3D subjective video quality tests. *EURASIP Journal on Image and Video Processing*, 2013, 50. <http://doi.org/10.1186/1687-5281-2013-50>
- [6] Kortum, P., & Sullivan, M. (2010). The Effect of Content Desirability on Subjective Video Quality Ratings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 52(1), 105–118. <http://doi.org/10.1177/0018720810366020>
- [7] Tarvainen, J., Westman, S., Oittinen, P. (2015). The Way Films Feel: Aesthetic Features and Mood in Film. *Psychology of Aesthetics, Creativity, and the Arts*. <http://doi.org/10.1037/a0039432>
- [8] ITU-R BT.500-12, Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union Recommendation, Rev. 12, 2009
- [9] ITU-T P.910, Subjective video quality assessment methods for multimedia applications, International Telecommunication Union Recommendation, 1999