



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE
WYDZIAŁ INFORMATYKI, ELEKTRONIKI I TELEKOMUNIKACJI

KATEDRA TELEKOMUNIKACJI

Praca dyplomowa magisterska

Opracowanie, analiza i ocena algorytmu wyznaczania wartości MOS
Develope an Algorithm Predicting MOS as a Function of FR Metrics

Autor:

Kierunek studiów:

Opiekun pracy:

Pola Łącz

Teleinformatyka

dr hab. Lucjan Janowski

Kraków, 2019

Uprzedzony o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.): „Kto przywłaszcza sobie autorstwo albo wprowadza w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystycznego wykonania albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także uprzedzony o odpowiedzialności dyscyplinarnej na podstawie art. 211 ust. 1 ustawy z dnia 27 lipca 2005 r. Prawo o szkolnictwie wyższym (t.j. Dz. U. z 2012 r. poz. 572, z późn. zm.): „Za naruszenie przepisów obowiązujących w uczelni oraz za czyny uchybiające godności studenta student ponosi odpowiedzialność dyscyplinarną przed komisją dyscyplinarną albo przed sądem koleżeńskim samorządu studenckiego, zwanym dalej «sądem koleżeńskim».”, oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i że nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.

Serdecznie dziękuję ... tu ciąg dalszych podziękowań np. dla promotora, żony, sąsiada itp.

Spis treści

1. Wstęp	7
2. Wprowadzenie teorytyczne	9
2.1. Wideo	9
2.2. Metryki	9
2.3. Algorytmy uczenia maszynowego	13
3. Metodologia badań	17
3.1. Dane	17
3.2. Modele	17
4. Analiza i wnioski	19
5. Podsumowanie	21

1. Wstęp

Pierwsze obrazy wideo powstały już na początku XX wieku i opierały się na mechanicznie obracających się dyskach. Technologia ta istniała głównie w sferze badań akademickich i nie zdominowała rynku. Dopiero z wprowadzeniem cathode-ray tube (CRT), wraz z telewizją analogową, wideo zaczęło być wykorzystywane komercyjnie. Z czasem rozwój technologii pozwolił na wprowadzenie telewizji cyfrowej, która zapewniała wyższą jakość obrazu oraz lepsze wykorzystanie zasobów. Wideo razem z audio okazały się również znakomitą środkiem wymiany informacji. Coraz częściej wykorzystywane do komunikacji w czasie rzeczywistym zastępując tradycyjne połączenie telefoniczne w biznesie oraz dla zwykłych użytkowników. Również rozwój na rynku telefonów wspomógł powszechność wideo. W momencie kiedy praktycznie każdy aparat zaczął posiadać kamerę, wideo zaczęło konkurować ze zdjęciami jako metoda na utrwalenia danej chwili. Codziennie tak rejestrowane obrazy są przekazywane do rodziny, znajomych oddalonych o tysiące kilometrów. Kolejnym przykładem kiedy wideo zastępuje tradycyjne formy przekazu są blogi internetowe do tej pory prowadzone na zasadzie artykułów/postów, teraz zaczęły wykorzystywać wideo jako metodę przekazu informacji.

Dzięki coraz większym przepustowości i szerokiemu dostępowi do Internetu w najnowszych czasach wykreował się jeszcze inny trend sprawiający że obrazy wideo są bardziej popularne. Mowa tu o platformach streamingowych takich jak - YouTube, Netflix czy HBOgo. Pozwalają one użytkownikom na oglądanie od krótkich filmików, przez serie, po pełnometrażowe filmy nawet w rozdzielczościach 4k. Przewidywane jest, że do 2022, aż 82 procent całego ruchu IP, to będzie wideo [1].

Wszystkie wymienione wyżej aspekty sprawiły, że wideo stało się codziennością w życiu większości ludzi.

Na obecnym etapie rozwoju technologii, oczekiwania odbiorcy co do jakości otrzymywanego wideo znacznie wzrosły. Na drugiej szali pozostają ograniczenia dotyczące medium i optymalnego wykorzystania zasobów po stronie klienta i serwera. Odnosząc się do powyższego istotną kwestią staje się monitorowanie jakości transmitowanego wideo i dostosowywanie go do potrzeb użytkownika. Jednak problem w ocenie jakości wideo jest tu o tyle trudny, że dotychczas najbardziej wiarygodnym wskaźnikiem jest tu opinia ludzka, nie powiązana (żadnym algorytmem?) z technicznymi aspektami obrazu?. W niniejszej Pracy zostanie przedstawiony algorytm pozwalający na bardziej zautomatyzowaną ocenę jakości wideo w oparciu o metryki full-reference (FR) i no-reference (NR) oraz zaprezentowana zostanie wykorzystana metodologia badań.

2. Wprowadzenie teorytyczne

W niniejszym rozdziale przedstawiono najważniejsze zagadnienia dla przeprowadzonych badań. Opisane one zostały w sposób pozwalający czytelnikowi na odpowiednie zrozumienie dalszej części pracy, pomijając niezwiązane szczegóły. Pierwsza część rozdziału dotyczy tematu wideo. Przedstawiono jego definicję, oraz wybrane cechy statystyczne biorące udział w trakcie badań. W kolejnej części przedstawiono zagadnienia z obszaru uczenia maszynowego. Wyjaśniona została jej ogólna koncepcja, a następnie opisano użyte algorytmy.

2.1. Wideo

<definicja ciężko znaleźć :(> Wideo jest formą elektronicznego zapisu sygnału wizji (analogowego bądź cyfrowego). W swojej surowej postaci jest to sekwencja pojedynczych ramek. Takie pliki zajmują bardzo dużo przestrzeni dyskowej. Przykładowo obraz o rozdzielczości 1920x1080 z 24 ramkami na sekundę o długości 30 sekund zajmuje ponad 20 GB w pamięci komputera. Tak duże rozmiary znacząco ograniczają możliwości przechowywania i transmisji dla zwykłych użytkowników. Dlatego praktycznie każdy plik wideo wykorzystuje kodeki, czyli pewne ustandaryzowane zasady kompresji/dekompresji. Do najpopularniejszych należą H.265 i H.264 powszechnie używany w Internecie do transmitowania multimediów [2]. Błędy podczas tego procesu, transmisji czy nagrywania mogą powodować powstanie zaburzeń w odtwarzanym obrazie. Aby mierzyć poziom uszkodzeń oraz monitorować stan wideo badacze zdefiniowali wiele metryk.

sprawdzić ile dok

2.2. Metryki

Jakość danego wideo jest najtrafniej oceniona dzięki tak zwanej ocenie subiektywnej. Polega ona na określeniu jakości odbioru przez człowieka na podstawie jego odczuć. Niestety aby takie badanie było miarodajne należy je przeprowadzić w odpowiednich warunkach, przygotować zestaw testów, zrekrutować uczestników oraz cały czas nadzorować jego przebieg. Wszystko to generuje duże koszty i zabiera cenny czas, ale przede wszystkim nie pozwala na ocenę jakości w czasie rzeczywistym na przykład podczas streamowania wideo. Problem ten rozwiązują metryki jakości wideo, które w przeciwieństwie do oceny subiektywnej, bazują na obiektywnych pomiarach właściwości[3].

Metryki video mogą zostać sklasyfikowane na podstawie tego czy do ich policzenia wymagana jest obecność niezakłóconego pliku video. Metryki, które wymagają takiego pliku określane są mianem *ang. full-reference* (FR). Porównują one właściwości zakłóconego i nie zakłóconego video, aby na podstawie tak otrzymanych informacji dokonać estymaty jakości. Innym rodzajem metryk są metryki *ang. no-reference* (NR). Dokonują one oceny jakości na podstawie zniekształconego video i nie wymagają do tego żadnych dodatkowych informacji[3].

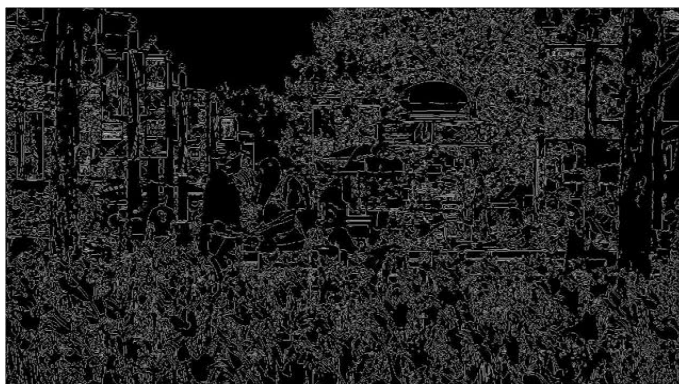
Poniżej zostały przedstawione wskaźniki jakości typu NR.

- **Blokowość**(*ang. Blockiness*) – powstaje podczas procesu kwantyzacji bloków pikseli i objawia się poprzez zauważalną granice między tymi blokami[4]. Skala przyjęta w pracy: 0-3570. Im większa wartość, tym mniej widoczne zakłócenie. Dla obrazu bez zakłóceń około 0.9-1.01[5]



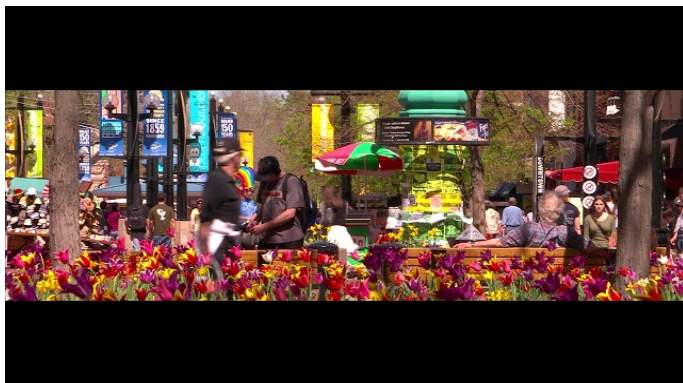
Rys. 2.1. Przykładowy obraz z zakłóceniami spowodowanymi blokowością[5]

- **Aktywność Przestrzenna** (*ang. Spatial Activity*) - opisuje położenie obiektu na ramce oraz jego relacje z innymi obiektami. Pozwala rozróżnić czy aktywność na danym wideo jest czynnością statyczną(osoba wykonująca czynność pozostaje w jednym miejscu) czy mobilną (osoba wykonująca czynność porusza się wzdłuż pola widzenia)[6]. Skala przyjęta w pracy: 0-270. Im większa wartość tym większa aktywność przestrzenna[5]



Rys. 2.2. Przykład detekcji obiektów w ramce, niezbędne do otrzymania informacji o aktywności przestrzennej[5]

- letterbox i pillarbox - letterbox jest techniką pozwalającą na wyświetlanie obrazów o wyższym współczynniku proporcji na odbiornikach o mniejszym współczynniku. Polega ona na dołożeniu dwóch czarnych pasów na górze i dole ramki.[7]. Natomiast pillarbox używane jest przy wyświetlaniu obrazu w mniejszym współczynniku proporcji na ekranie o większym. Do obrazu dodawane są wtedy pasy po obu bokach ekranu. Skala przyjęta w pracy 0-1.



(a) opis[5]



(b) opis[5]

- Straty bloków (*ang. blockloss*) - wskaźnik informujący o ilości brakujących bloków(używanych podczas DCT(Dyskretna transformacja kosinusowa)). Skala przyjęta w pracy: 0-200.Im większa wartość tym większe zniekształcenie obrazu.

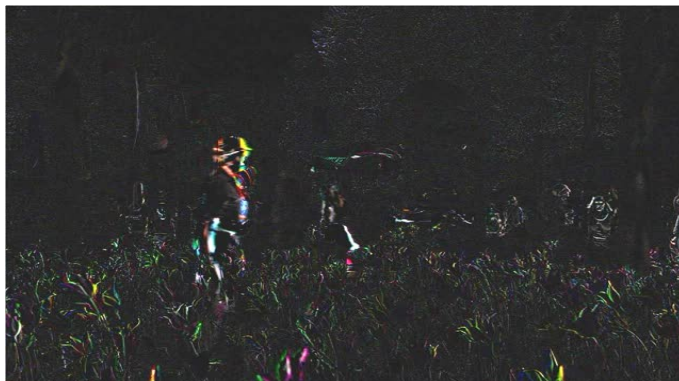


Rys. 2.4. opis[5]

- Rozmycie (*ang. blur*) - Istnieje wiele rodzajów rozmycia, między innymi te niepożądane, występujące po przeprowadzeniu kompresji bądź przez poruszenie kamery podczas nagrywania. Gdy występuje ten efekt obiekty na obrazie tracą ostrość krawędzi.[8]. Skala przyjęta w pracy: 0-70.
- Aktywność czasowa(*ang. Temporal activity*) - opisuje intensywność ruchu obiektów w czasie. Często łączone z wskaźnikiem Aktywność przestrzenna. Skala przyjęta w pracy: 0-255.
- blackout - opisuje zjawisko kiedy obraz całkowicie zanika[9]. Skala przyjęta w pracy : 0-1



Rys. 2.5. opis[5]



Rys. 2.6. opis[5]

- Zamrożenie (*ang. freezing*) - Informuje o efekcie czasowego zatrzymania zatrzymania obrazu, które powoduje odczucie "przycięcia" wideo[10]. Skala przyjęta w pracy: 0-1.
- Ekspozycja (*ang. exposure*) - ten typ zakłóceń jest spowodowany niezbilanowaną jasnością ramek. Widz ma odczucie ze wideo jest zbyt ciemna lub zbyt jasne[10]. Skala przyjęta w badaniach 0-255.

- contrast
- brightness
- interlace
- noise
- slice
- flickering

metryki FR i ich wymienienie w oparciu o <https://ieeexplore.ieee.org/document/5506331>

- SSIM



Rys. 2.7. opis[5]



Rys. 2.8. opis[5]

- VMAF
- PSNR
- itp

pozostałe cechy video:

- Rozdzielczość – miara określająca rozmiar ramki. Jednostką są pixele. Podawana jest zazwyczaj w następujący sposób: szerokość x wysokość. Do badań zostały użyto wideo o rozdzielczościach: 3840x2160, 1920x1080, 704x576, 640x480, 352x288.
- Klatki na sekundę(ang.frames per second, fps) – liczba ramek wyświetlonych w czasie sekundy. W telewizji jest to 25 ramek na sekundę. Do badań użyto: 60, 30, 25, 24 fps

2.3. Algorytmy uczenia maszynowego

- Ogólne informacje uczeniu maszynowym/
- Przedstawienie wybranych algorytmów



Rys. 2.9. opis[5]



Rys. 2.10. opis[5]



Rys. 2.11. opis[5]



Rys. 2.12. opis[5]



Rys. 2.13. opis[5]

3. Metodologia badań

3.1. Dane

- Wybrane narzędzia
- Opis zebranych danych
- Przedstawienie data flow(pobieranie-> czyszczenie->normalizacja->przygotowanie formatu dla modeli).
- Wizualizacja danych

3.2. Modele

- Opis zastosowanych parametrów/technik podczas trenowania.
- Przedstawienie wyników

4. Analiza i wnioski

- Interpretacja wyników
- Opis innych czynników mogących zaburzyć ich prawdziwość
- Co nie zostało uwzględnione

5. Podsumowanie

- Czy cel pracy został osiągnięty.
- Możliwości rozbudowy

Bibliografia

- [1] Cisco. *Forecast and Trends 2017-2022*.
<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual>
2019.
- [2] Anthony Romero. *On Line Alvis Manual*.
<https://video.ibm.com/blog/streaming-video-tips/what-is-video-encoding-coo>
2018.
- [3] Christian Keimel. *Design of Video Quality Metrics with Multi-Way Data Analysis: A data driven approach*. Spraw. tech. 2016, s. 31–32.
- [4] M. Ghanbari M.T. Qadri K.T. Tan. *Frequency domain blockiness measurement for image quality assessment, Introduction*.
<https://ieeexplore.ieee.org/document/5645871>. 2010.
- [5] Katedra Telekomunikacji AGH IEiT. *Video Quality Indicators*.
<http://vq.kt.agh.edu.pl/index.html>. 9999.
- [6] William Blackburn Shreeya Sengupta Hui Wang. *Spatial information in classification of activity videos*. Spraw. tech. School of Computing i Mathematics, University Of Ulster, 2015.
- [7] *16:9 aspect ratio conversion by letterbox method for an MPEG image*. 2000.
- [8] S. Winkler P. Marziliano F. Dufaux. *A no-reference perceptual blur metric*.
<https://ieeexplore.ieee.org/document/5645871>. 2002.
- [9] Abdelhamid Mellouki Hala Rifaï Samer Mohammed. *A brief synthesis of QoS-QoE methodologies*.
<https://ieeexplore.ieee.org/document/5898880>. 2011.
- [10] Mylène C. Q. Farias Mikołaj Leszczuk Mateusz Hanusiak. *Recent developments in visual quality monitoring by key performance indicators*. Video artefacts and related key performance indicators for automated quality checking. 2014.