

Filtracja sygnału w metodach ciągłej oceny błędów kompresji materiału filmowego na rzecz automatyzacji procesu monitorowania jakości

▶ Anna Ostaszewska
Sabina Żebrowska-Łucyk

Skutkiem ubocznym kompresji stratnej stosowanej w przypadku materiału filmowego jest pojawienie się widocznych zniekształceń obrazu. Zniekształcenia te nazywane są błędami kompresji, a stopień ich nasilenia decyduje o jakości produktu. Najlepszą metodą kontroli jakości jest poddanie materiału badaniom z udziałem grupy przeszkolonych obserwatorów, jednak nie zawsze jest to możliwe, np. w przypadku całodobowej kontroli nadawanej cyfrowo telewizji. Wówczas człowieka zastępuje analizator – algorytm pozwalający na automatyczną ocenę jakości kompresji. Analizatory tworzone są na podstawie wyników uzyskanych na drodze badań z udziałem grupy obserwatorów. Wyniki te charakteryzują się dużymi przedziałami ufności uniemożliwiającymi wnioskowanie statystyczne i projektowanie analizatorów. Artykuł prezentuje modyfikację metody prowadzenia badań percepcyjnych oraz sposób filtracji, który pozwala na minimalizację przedziałów ufności, co może przyczynić się do rozwoju analizatorów.

Dynamiczny rozwój telewizji (w tym wysokiej rozdzielczości) nadawanej cyfrowo lub przez Internet, telekonferencji oraz wszelkiego rodzaju multimediów spotęgował zapotrzebowanie na efektywne techniki kompresji stratnej, w wyniku której materiał zakodowany zajmuje na nośniku mniej miejsca niż materiał źródłowy, kosztem utraty pewnej części informacji. Różnice pomiędzy materiałem po kompresji a materiałem oryginalnym mogą skutkować dostrzegalnym pogorszeniem jakości o nasileniu zależnym od typu kodera, parametrów kompresji, specyfiki materiału i innych czynników.

Postęp w dziedzinie technik kompresji obrazów ruchomych jest uwarunkowany rozwojem metod oceny jakości produktów kompresji. Zasadnicze znaczenie mają metody zwane percepcyjnymi (in. subiektywnymi), które polegają na prezentowaniu widzom materiału filmowego uprzednio poddanego kompresji, rejestracji ocen przyznanych przez obserwatorów, a następnie przetworzeniu i analizie danych. Wyniki tego typu pomiarów służą nie tylko do ostatecznej oceny nowych technik kompresji, wyboru parametrów kodowania, oceny jakości transmisji sygnału TV, ale również do tworzenia oraz kalibracji analizatorów – algorytmów pozwalających na automatyczną ocenę jakości kompresji.

Metody percepcyjnej oceny jakości kompresji

Metody percepcyjne można podzielić na punktowe (obserwator przyznaje jedną ocenę po obejrzeniu kilkunastosekundowej sekwencji) oraz ciągłe w czasie (obserwator nieprzerwanie na bieżąco podaje oceny jakości trwającego do 30 min filmu, posługując się urządzeniem dołączonym do komputera). Prace skoncentrowano na metodach ciągłych, które pozwalają na uzyskanie większej liczby punktów pomiarowych, co ma istotne znaczenie w przypadku tworzenia i kalibracji analizatorów – algorytmów naśladujących zachowanie obserwatora.

Do najważniejszych metod ciągłych należą zalecane przez International Telecommunication Union: jednobodźcowa – *Single Stimulus Continuous Quality Evaluation* (SSCQE) [1] oraz dwubodźcowa – *Simultaneous Double Stimulus for Continuous Evaluation* (SDSCE) [2]. Metoda SSCQE charakteryzuje się brakiem materiału odniesienia; obserwatorzy oceniają wyświetlany film, nie znając materiału źródłowego. W metodzie SDSCE obserwatorzy oglądają zsynchronizowane materiały: jeden skompresowany, drugi źródłowy. Pomimo stosowania metod ciągłych przez różne ośrodki badawcze na świecie, metody te są słabo opisane. Brak jest doniesień zarówno na temat budowy i oprogramowania stanowisk badawczych, jak i sposobu przetwarzania danych pomiarowych. Literatura informuje głównie o badaniach wpływu materiału (długości sesji, treści oglądanych sekwencji, towarzyszącego dźwięku) na przebieg sygnału oceny w cza-

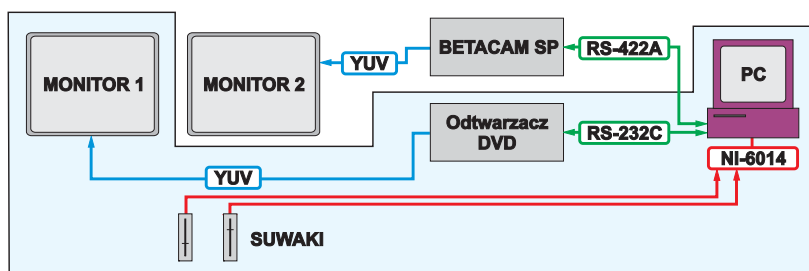
▶ mgr inż. Anna Ostaszewska,
dr hab. inż. Sabina Żebrowska-Łucyk, prof. nzw. PW
Instytut Metrologii i Inżynierii Biomedycznej
Politechnika Warszawska

nie oraz o właściwościach specyficznego przyrządu pomiarowego, jakim jest obserwator-arbiter, w tym o ograniczonej pojemności pamięci, roli doświadczenia w ocenie jakości skompresowanych materiałów filmowych oraz o wpływie opisu skali ocen na ich rozkład. Zalecenia ITU wskazują na potrzebę krytycznego odnoszenia się do pozyskanych danych pomiarowych i przedstawiają metodę odrzucania ocen niespójnych.

Badania wstępne przeprowadzone w Instytucie Metrologii i Inżynierii Biomedycznej wykazały, że duże rozrzuty ocen uniemożliwiają statystyczne wnioskowanie na temat wpływu na rozkład ocen poszczególnych czynników, takich jak m.in. zastosowanie lub brak wzorca odniesienia. Nie pozwala to na racjonalny wybór metody ani ocenę skutków ewentualnych jej modyfikacji. W związku z tym podjęto prace mające na celu zminimalizowanie przedziałów ufności wybranych parametrów statystycznych i zbadanie wpływu wzorca odniesienia na rozkład ocen. Zastosowano metodę postępowania zbliżoną do używanych w metrologii, zwłaszcza w tych jej obszarach, które charakteryzują się dużym poziomem zakłóceń i dominującym wpływem obserwatora. W ramach prac zbadano obie wspomniane metody oceny jakości kompresji materiału filmowego (SSCQE i SDSCE) w celu ustalenia ich właściwości i ich porównania. Wskazano również możliwości udoskonalenia wymienionych metod pod kątem pozyskiwania danych do tworzenia i kalibracji analizatorów jakości kompresji materiału filmowego.

Stanowisko pomiarowe

Opracowane i zrealizowane w IMiIB stanowisko badawcze (jako pierwsze w Polsce) umożliwiło prowadzenie badań metodami SSCQE oraz SDSCE (rys. 1). Do odtworzenia materiału testowego użyto odtwarzacza Pioneer DVD-V7300D, zapewniającego wysoki stosunek sygnału do szumu (60 dB) oraz magnetowidu studyjnego SONY UVW-1800 BETACAM SP. Odtwarzacze komunikują się z komputerem przez porty RS. Do prezentacji materiałów posłużyły dwa 20-calowe monitory Trinitron® Color Video Monitor SONY PVM-20M4E, wyświetlające 800 linii, o stabilizowanym balansie bieli. Urządzenie do wprowadzania oceny ma postać potencjometru suwakowego, z podziałką o zakresie od 0 do 100. Sygnałem mierzonym jest napięcie podawane do umieszczonej w komputerze karty pomiarowej National Instruments NI-6014 o rozdzielczości 16 bitów. Komponentowy sygnał Y/R-Y/B-Y z odtwarzaczy podawany jest do monitorów. Oprogramowanie stanowiska utworzono w środowisku LabVIEW. Aplikacja konfiguruje kartę pomiarową, obsługuje porty RS, linearyzuje charakterystykę przetwarzania potencjometrów, próbkuje sygnał pomiarowy z częstotliwością 2 Hz, zapisuje na dysku pobierane dane, umożliwia zdalne sterowanie odtwarzaczem DVD oraz monitoruje przy-



Rys. 1. Schemat stanowiska pomiarowego do badań jakości kompresji metodami SSCQE i SDSCE; poza ramką elementy zbędne w badaniach metodą SSCQE

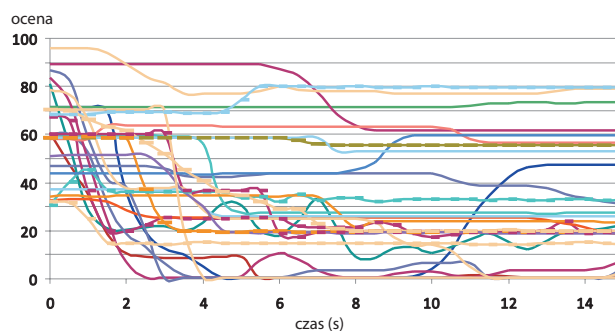
chodzące sygnały. Komunikacja odtwarzaczy z komputerem pozwala na zachowanie pełnej synchronizacji wyświetlanych obrazów i sygnału z potencjometrów z dokładnością do pojedynczej ramki. Podczas badań metodą SSCQE używany jest jedynie monitor 1, na którym wyświetlany jest materiał poddany uprzednio kompresji, odtwarzany z DVD. Na potrzeby SDSCE na monitorze 2 wyświetlany jest dodatkowo materiał referencyjny z magnetowidu.

Badano wpływ trzech czynników na ocenę jakości materiału: wartość strumienia bitowego (*bitrate*), liczbę ramek typu B w strukturze GOP (*Group of Pictures*) oraz zawartość (treść) sceny. Przygotowano zbiór czterech 15-sekundowych sekwencji zakodowanych w standardzie MPEG-2 na 10 poziomach strumienia bitowego (w przedziale od 2 Mbps do 5 Mbps), przy wszystkich trzech możliwych strukturach 13-obrazowej GOP: bez ramek typu B, z jedną i dwiema. Powstała biblioteka 120 komponentów, które służą do realizacji wybranych planów eksperymentu przy użyciu obu metod.

Wyniki badań i propozycja analizy danych

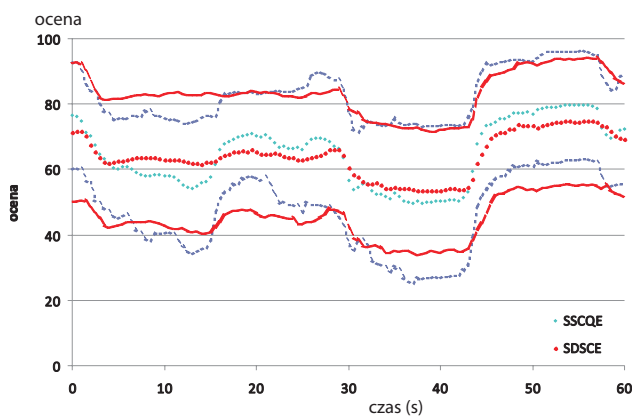
Przeprowadzono serię eksperymentów metodami SSCQE i SDSCE pozyskując łącznie blisko 300 tys. wyników pomiarów.

Przetworzone zgodnie z zaleceniami ITU [1, 2] wyniki badań SSCQE i SDSCE charakteryzują się znacznym rozrzutem spowodowanym różnym czasem reakcji poszczególnych obserwatorów, różną rozpiętością wykorzystanej skali oraz różną percepcją pojawiających się błędów kompresji. Zalecane przez ITU meto-



Rys. 2. Fragment zbioru surowych sygnałów pomiarowych pozyskanych od 45 obserwatorów metodą SSCQE zgodnie z [1]

dy filtracji danych pomiarowych w celu odrzucenia wyników niewiarygodnych [2] nie przyniosły żadnej poprawy. Ustalono, że odchylenie standardowe ocen dla poszczególnych wartości strumienia wynosi aż 14–18 % całkowitego zakresu (rys. 3). Przyjmując za poziom istotności $\alpha = 0,05$, nie stwierdzono podstaw statystycznych do odrzucenia hipotezy zerowej o równości wartości oczekiwanych sygnałów pozyskanych różnymi metodami, jednak w związku z dużą wariancją ocen ryzyko popełnienia błędu drugiego rodzaju jest bardzo wysokie (test ma słabą moc). Co więcej, używanie tak silnie zaszumionych wyników badań do kalibracji analizatorów sprzętowych, doboru parametrów kompresji lub wnioskowania na temat właściwości ludzkiej percepcji na rzecz tworzenia jej modelu grozi formułowaniem nieuprawnionych wniosków.



Rys. 3. Oceny średnie (MOS) i przedziały o szerokości odchylenia standardowego, wyznaczone metodą SSCQE (19 obserwatorów) i SDSCE (57 obserwatorów) zgodnie z [1]

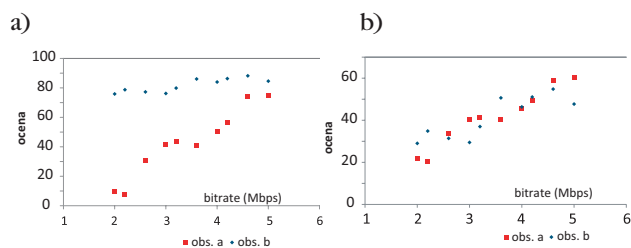
W tej sytuacji priorytetem staje się dążenie do zawężenia przedziału ufności dla wartości oczekiwanych przy ustalonym poziomie ufności. Istnieją dwie koncepcje: a) zwiększenie liczby obserwacji (kosztowne i czasochłonne) oraz b) przefiltrowanie danych surowych w celu usunięcia danych niespójnych, połączone z odpowiednim przetworzeniem pozostałych wyników pomiaru.

Koncentrując się na drugim podejściu, zmodyfikowano metodę prowadzenia badania. Skrócono czas trwania poszczególnych scen do 15 s, dzięki czemu w ciągu eksperymentu ograniczonego zaleceniami do 30 min [1, 2], każdy z obserwatorów dwukrotnie ocenił ten sam materiał. Tego typu modyfikacja pozwoliła na opracowanie nowej metody filtracji pozyskanego od obserwatorów sygnału. Filtracja ta opiera się na założeniu, że wartości średnie i zakres skali używanej przez poszczególnych obserwatorów mogą się znacznie różnić, ale oczekuje się zgodności rang przypisanych przez poszczególnych obserwatorów z rangami ocen średnich. Ponadto oceny wprowadzone przez poszczególnych obserwatorów przy obserwacji tego samego materiału nie powinny się zbytnio różnić, jednak można oczekiwać chwilowej dekoncentracji podczas długiej sesji testowej nawet u uważnych obserwatorów i takie fragmenty sygnałów powinny zostać usunięte.

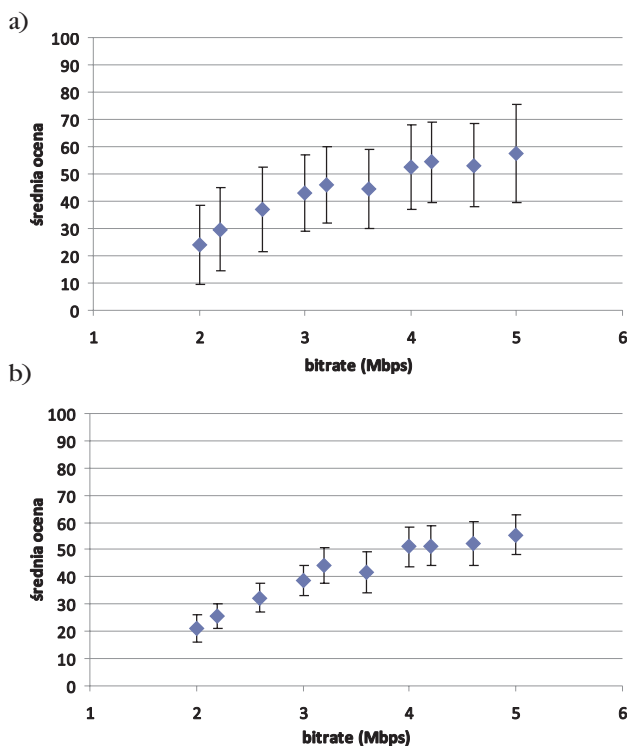
Koncepcja filtracji zasadza się na wyodrębnieniu dwóch rodzajów niespójności danych: *wewnętrznej* – niezgodności ocen przyznanych przez danego obserwatora temu samemu materiałowi oglądanemu dwukrotnie oraz niespójności *zewnętrznej* – braku zgodności ocen obserwatora z ogólną średnią ocen.

Opracowana metoda filtracji danych pomiarowych stanowi sekwencję operacji, które niwelują różnice czasu reakcji obserwatora, standaryzują sygnały (rys. 4) oraz pozwalają na obiektywną eliminację fragmentów sygnałów (lub całych sygnałów) niespójnych wewnętrznie lub zewnętrznie. Dobierając kryteria usuwania danych i testy statystyczne wzięto pod uwagę porządkowy (a nie ilorazowy czy przedziałowy) charakter skali pomiarowej oraz rozkład danych pomiarowych, odbiegający od gaussowskiego.

Operacjom takim poddano wyniki pomiarów uzyskane metodą SSCQE z udziałem 45 osób, z których każda dwukrotnie oceniła ten sam materiał zakodowany przy 10 wartościach strumienia bitowego. W wy-



Rys. 4. Oceny dwóch obserwatorów uśrednione w czasie dla materiału zakodowanego przy ustalonej wartości strumienia bitowego: a) obliczone na podstawie danych surowych, b) po standaryzacji



Rys. 5. Oceny średnie (MOS) i odchylenie standardowe ocen wyznaczone metodą SSCQE (45 obserwatorów, scena *mobile*) a) na podstawie danych surowych, b) po operacjach filtracji i standaryzacji danych

niku analizy danych z zastosowaniem współczynnika korelacji rang Spearmana wykluczono z dalszych badań sygnały pochodzące od 17 % obserwatorów. Ponadto, stosując statystyki k-Mandela i h-Mandela, odrzucono niespójne fragmenty ocen zaobserwowane w części sygnałów na niektórych poziomach kodowania. Cały ciąg zaproponowanych i sprawdzonych empirycznie operacji pozwolił na dwukrotne zmniejszenie odchylenia standardowego ocen (rys. 5).

Podsumowanie

W wyniku zastosowania opracowanej metody filtracji eliminowane są nie tylko całe sygnały od niektórych obserwatorów, ale również fragmenty wyników uzyskanych podczas epizodów braku koncentracji, jakie mogą zdarzyć się nawet doświadczonym arbitrom. Tego typu podejście stanowi nowość w dziedzinie badań jakości wizualnej i pozwala na znaczne zawężenie przedziałów ufności wyników realizowanych pomiarów.

Plan dalszych prac obejmuje wykonanie analogicznych operacji na sygnale ciągłym w czasie, przeprowadzenie opisanych zabiegów na wynikach pomiarów pozyskanych metodą SDSCE oraz porównanie wyników uzyskanych obiema metodami i wyników z analizatora sprzętowego generującego ocenę ciągłą w czasie.

Końcowym efektem opisanych prac będzie przedstawienie modyfikacji, które umożliwią wyznaczenie i poprawę dokładności metod ciągłej oceny błędów kompresji, co powinno przyczynić się do rozwoju analizatorów jakości wyznaczających oceny zgodne z odbiorem percepcyjnym.

Praca naukowa finansowana ze środków na naukę w latach 2007-2009 jako projekt badawczy.

Bibliografia

1. ITU-T Recommendation P.911 (1996), *Subjective audiovisual quality assessment methods for multimedia applications*.
2. ITU-R Recommendation BT.500-11 (1998), *Methodology for the subjective assessment of the Quality of Television Pictures*.
3. Alpert Th., Evain J.-P.: *Subjective quality evaluation – The SSCQE and DSCQE methodologies*, EBU Technical Review Spring 1997, pp. 21–30.
4. Pinson M., Wolf S.: *Comparing subjective video quality testing methodologies*, SPIE Video Communications and Image Processing Conference, Lugano, Switzerland, Jul. 8–11 2003, vol. 5150 (3), pp. 573–582.
5. Ostaszewska A., Żebrowska-Łucyk S., Kłoda R.: *Metrology properties of human observer in compressed video quality evaluation*, XVIII IMEKO WORLD CONGRESS, Metrology for a Sustainable Development Rio de Janeiro, Brazil 2006.

