

ULEPSZONE STEROWANIE KODEREM JPEG Z RDOQ ENHANCED CONTROL IN JPEG ENCODER WITH RDOQ

Jakub Stankowski, Kacper Winiecki, Krzysztof Klimaszewski, Tomasz Grajek

Instytut Telekomunikacji Multimedialnej, Politechnika Poznańska, Poznań
{jakub.stankowski, krzysztof.klimaszewski, tomasz.grajek}@put.poznan.pl

Streszczenie: W pracy przedstawiono dwa modele użyteczne do szybkiego wyznaczania wartości mnożnika Lagrange'a w koderze JPEG wykorzystującym technikę RDOQ. Z wykorzystaniem licznego zestawu obrazów testowych zmierzono zostały przyspieszenie procesu kodowania dla różnych warunków pracy kodera oraz wpływ zastosowanego modelu na sprawność kompresji. Uzyskano nawet 2-krotne przyspieszenie pracy kodera przy pomijalnie małym wpływie na efektywność kompresji.

Abstract: This paper presents two models useful for the fast determination of the Lagrange multiplier in a JPEG encoder using the RDOQ technique. Using a large set of test images, the acceleration of the encoding process was measured for various encoder operating conditions, as well as the impact of the model used on compression efficiency. An acceleration of up to twofold was achieved, with a negligible impact on compression efficiency.

Słowa kluczowe: JPEG, kompresja obrazu, mnożnik Lagrange'a, RDOQ.

Keywords: image compression, JPEG, Lagrange multiplier, RDOQ.

1. WSTĘP

Pomimo faktu, że format kompresji obrazu JPEG, opisany w normie ITU-T T.81 został opublikowany w roku 1991, a od tego czasu opracowane zostały znacznie bardziej wydajne formaty kompresji obrazów, to JPEG wciąż pozostaje dominującym sposobem stratnej kompresji obrazów [4][5][13]. Kodery JPEG podlegają ciągłemu rozwojowi [12], zachowując kompatybilność z formatem opisanym w normie [8]. Rozwój koderów pozwala na poprawę sprawności kodowania poprzez uzyskanie lepszej jakości zrekonstruowanego obrazu przy niezmienionej liczbie bitów lub zmniejszenie liczby bitów przy tej samej jakości. Osobnym kierunkiem rozwoju koderów może być skrócenie czasu kodowania obrazu.

W [7] opisano koder JPEG wykorzystujący technikę Rate-Distortion Optimized Quantization (RDOQ) do poprawy sprawności kompresji. W niniejszej pracy zaprezentowana została modyfikacja tego kodera. Proponowane usprawnienie polega na opracowaniu implementacji wzorów przyspieszających wyznaczenie wartości parametru λ (mnożnika Lagrange'a) wykorzystywanego w procesie optymalizacji wykonywanym przez koder.

W artykule zaprezentowane i porównane są dwie metody wyznaczania wartości parametru λ , bazujące na jednej lub na trzech zmiennych.

2. OPTIMALIZACJA W KODERZE JPEG

W tradycyjnych implementacjach kodera JPEG współczynniki transformaty podlegają kwantowaniu przy użyciu określonej tabeli kwantyzacji, przeskalowanej o parametr jakości Q , po czym skwantowane współczynniki zapisywane są w strumieniu. Wykorzystywana w opisywanej pracy metoda optymalizacji polega na zmianie wartości skwantowanych współczynników transformaty. Zmiany te prowadzą do zmniejszenia liczby bitów zapisywanych do strumienia kosztem zmiany jakości zrekonstruowanego bloku. W praktyce najbardziej wydajną metodą optymalizacji jest usuwanie współczynników o małych wartościach odpowiadających wysokim częstotliwościom przestrzennym w bloku lub zmiana wartości takich współczynników o ± 1 . W procesie decyzyjnym wykorzystywany jest wzór (1) określający koszt kodowania bloku, który podlega minimalizacji.

$$J = D + \lambda \cdot R \quad (1)$$

We wzorze (1) J oznacza koszt kodowania bloku, D to miara zniekształceń zrekonstruowanego bloku (na przykład MSE – Mean Squared Error), R to liczba bitów zużytych na zakodowanie bloku, natomiast λ to mnożnik Lagrange'a, którego odpowiednio dobrana wartość pozwala na uzyskanie optymalnego sposobu kodowania danego bloku lub całego obrazu. W oprogramowaniu opisanym w [7] wartość współczynnika λ wyznaczana została na podstawie pomiaru wartości pochodnej:

$$\lambda = -\frac{dD}{dR} \quad (2)$$

Pomiar ten był dokonywany dla ustalonej wartości współczynnika sterującego jakością kompresji Q poprzez wykonanie pomocniczego zakodowania tego samego obrazu z użyciem wartości współczynnika $Q + 2$ oraz $Q - 2$, a następnie obliczenia pochodnej na podstawie uzyskanych wartości D i R dla obu kodowań. Metoda ta pozwala na uzyskanie poprawnych wartości λ , jednak ze względu na konieczność dodatkowego dwukrotnego wykonania częściowej kompresji obrazu jest czasochłonna i może stanowić nawet 60% czasu kodowania (por. Tabela 3). Obserwacja ta była motywacją do opracowania szybkiej metody wyznaczania wartości parametru λ .

3. OPIS PRAC

Na potrzeby opracowania szybszej metody wyznaczenia współczynnika λ zebrane zostały dane z procesu kodowania różnorodnych obrazów z bazy sekwencji testowych utworzonej między innymi z [1]. W bazie znalazło się łącznie 5967 obrazów pochodzących z szesnastu różnorodnych sekwencji testowych o rozdzielczości 1920x1080. Każdy obraz z bazy podlegał zakodowaniu z użyciem pełnego zakresu wartości parametru sterującego jakością $Q \in \langle 0, 100 \rangle$. Dla każdego kodowania zebrano wartości R i D . Wyniki i parametry kodowania każdego obrazu zapisane zostały w utworzonej bazie danych. Użyty do tego kodowania koder nie przeprowadzał optymalizacji wartości skwantowanych współczynników transformaty. Odrębne zestawy wartości wyznaczone zostały dla kodowania luminancji i chrominancji. W badaniach wykorzystano typowe tablice kwantowania (tj. bazujące na tablicy określonej w ITU-T T.81 przeskalowane w zależności od użytego parametru Q). W trakcie pomiarów w sumie wykonano kompresję i zebrano dane dla $5967 \cdot 101 = 602667$ obrazów.

Na podstawie zebranych danych opracowane zostały uniwersalne wzory pozwalające na wyznaczenie wartości λ bez konieczności przeprowadzania pomocniczych kodowań. Działania skierowane zostały na uzyskanie dwóch różnych modeli – prostszego i szybszego oraz bardziej złożonego, dającego potencjalnie lepsze wyniki kompresji.

3.1. Model z jedną zmienną

Przeprowadzone zostały prace pozwalające na wyznaczenie modelu z jedną zmienną, czyli zależności optymalnej wartości parametru λ od wartości parametru Q . Modele opracowano z wykorzystaniem metod dobierania wzorów empirycznych opisanych w [3].

Pierwszym krokiem było wyznaczenie funkcji przybliżającej zależność $D(R)$ – miara zniekształceń obrazu D w funkcji liczby bitów na obraz R dla każdego zbadanego obrazu oddzielnie. Jako miara zniekształceń wykorzystana została miara *Sum of Squared Differences* (SSD) – suma kwadratów różnic próbek obrazu oryginalnego i zrekonstruowanego. Dla luminancji przyjęto przybliżenie modelem (3)

$$D_L(R_L) = a_L \cdot R_L^b + c \cdot R_L^d \quad (3)$$

natomiast dla chrominancji, ze względu na bardziej złożoną zależność, przyjęto przybliżenie modelem (4).

$$D_C(R_C) = e^{a \cdot R_C^b + c \cdot R_C^d} \quad (4)$$

Przybliżanie wykorzystywało metodę najmniejszych kwadratów [3]. Na podstawie wyznaczonych przybliżeń (wzory 3 i 4) wyznaczono analitycznie wartości λ na podstawie (2). W następnym kroku określono zależność λ od wartości parametru Q dla każdego z badanych obrazów z wykorzystaniem zapisanej w bazie danych bezpośredniej zależności R od Q dla każdego z kodowanych obrazów. Kolejno wyznaczono zagregowaną po wszystkich zakodowanych obrazach zależność $\lambda(Q)$ jako

medianę zależności dla wszystkich obrazów, wygładzoną filtrem Savitzkiego-Golaya [9]. Ostatnim krokiem było odnalezienie analitycznej zależności opisanej wzorem (5) z wykorzystaniem metody najmniejszych kwadratów dla luminancji i chrominancji.

$$\ln(\lambda) \approx a \cdot \ln^3(Q) + b \cdot \ln^2(Q) + c \cdot \ln(Q) + d \quad (5)$$

Uzyskane wartości wykorzystane zostały w zmodyfikowanym koderze JPEG i przedstawione są w Tabeli 1.

Tabela 1. Wartości parametrów modelu z jedną zmienną (5) dla luminancji (Y) i chrominancji (Cb, Cr)

Kanał	a	b	c	d
Y	-0,6203	6,352	-23,21	33,71
Cb, Cr	0,4974	5,812	-24,52	40,42

3.2. Model z trzema zmiennymi

W kolejnym etapie, przeprowadzone zostały prace pozwalające na wyznaczenie modelu z trzema zmiennymi, czyli zależności optymalnej wartości parametru λ od wartości parametrów Q , BPP i MSE. Zmienna BPP to liczba bitów na piksel wyznaczona na podstawie wstępnego, szybkiego kodowania obrazu bez wykorzystania optymalizacji RDOQ, natomiast MSE to średnia wartość SSD na piksel obrazu, wyznaczona podczas tego samego szybkiego kodowania obrazu. Jak zatem widać, metoda z 3 zmiennymi wymaga dodatkowego kodowania obrazu, co zmniejsza potencjalne przyspieszenie koder, jednak daje lepsze wyniki kodowania. W modelu z trzema zmiennymi wartość współczynnika λ jest przybliżana wzorem (6).

$$\begin{aligned} \ln(\lambda) \approx & a + b \cdot \ln(Q) + c \cdot \ln(BPP) + \\ & + d \cdot \ln(MSE) + e \cdot \ln(Q) \cdot \ln(BPP) + \\ & + f \cdot \ln^2(Q) + g \cdot \ln^2(BPP) \end{aligned} \quad (6)$$

W procesie wyznaczania współczynników wzoru (6) wykorzystano algorytm dopasowania metodą *Random Sample Consensus* (RANSAC) [6], co podyktowane było faktem występowania dużej liczby punktów pomiarowych wyraźnie odbiegających od większości. Algorytm RANSAC działa na pełnym zbiorze danych zebranych w czasie pomiarów. Nie zostały tu przeprowadzone kroki pośrednie, jak w metodzie z jedną zmienną, ze względu na większy stopień skomplikowania zależności oraz duże rozpiętości wartości współczynników pomiędzy różnymi obrazami i trudności w zagregowaniu wyników kolejnych etapów. Wartości parametrów modelu (6) zostały przedstawione w Tabeli 2.

Tabela 2. Wartości parametrów modelu z trzema zmiennymi (6) dla luminancji (Y) i chrominancji (Cb, Cr)

Kanał	a	b	c	d	e	f	g
Y	2,580	-0,3591	-0,6348	0,7703	-0,0333	-0,0295	0,1913
Cb, Cr	3,069	0,5248	-0,4314	0,6414	-0,0739	-0,2562	0

4. WERYFIKACJA EKSPERYMENTALNA

W celu weryfikacji skuteczności zaproponowanych modeli, odpowiednio zmodyfikowany koder JPEG [11] poddano testom wydajnościowym oraz jakościowym. Badaniom podlegały czas wykonywania poszczególnych etapów kodowania oraz wpływ wprowadzonych uproszczeń na końcową jakość obrazu, wyrażoną w mierze Bjøntegaard Delta-Rate (dRate) [2].

4.1. Redukcja czasu wyznaczania wartości λ

Zasadniczym celem optymalizacji było przyspieszenie procesu wyznaczania parametru λ (mnożnika Lagrange'a), który w standardowym wariancie RDOQ jest wysoce czasochłonny z uwagi na konieczność przeprowadzania dwóch pomocniczych kodowań. Jak zaprezentowano w Tabeli 3, dla wariantu kodowania o wysokiej jakości ($Q=90$) przykładowej sekwencji „pedestrian_area”, czas wyznaczania mnożnika w modelu odniesienia wynosi 2971,53 μ s. Zastosowanie modelu z jedną zmienną, jak wskazują dane zaprezentowane w Tabeli 4, redukuje ten czas do wartości 0,53 μ s, zatem w praktyce zaniedbywalnego. Z kolei model oparty na trzech zmiennych wymaga jednego pomocniczego kodowania i pozwala na uzyskanie czasu na poziomie 880,81 μ s, dając ponad trzykrotną redukcję czasu względem odniesienia.

Tabela 3. Przykładowe średnie czasy poszczególnych etapów kodowania JPEG dla sekwencji „pedestrian_area”

Nazwa etapu przetwarzania	Czas przetwarzania [μ s]		
	Q=90	Q=60	Q=30
Transformacja	307,65	309,58	312,17
Kwantyzacja i szeregowanie	291,28	294,24	294,58
Wyznaczanie λ	2971,53	2776,89	2781,41
RDOQ	5397,57	1523,21	816,71
Optymalizacja tablic Huffmana (HTO)	276,99	251,17	249,84
Kodowanie entropijne	242,41	124,98	102,89
Zapis markerów i tabel (DQT i DHT)	0,67	0,79	0,79
Zapis obrazu	16,23	9,47	7,34
Kodowanie obrazu (w sumie)	9488,11	5280,87	4558,41

Tabela 4. Przykładowe średnie czasy wyznaczania parametru λ dla sekwencji „pedestrian_area” z wykorzystaniem zaproponowanych modeli

Model	Czas przetwarzania [μ s]		
	Q=90	Q=60	Q=30
odniesienia	2971,53	2776,89	2781,41
1 zmienna	0,53	0,48	0,50
3 zmienne	880,81	804,38	807,11

Analiza średnich czasów dla sekwencji testowej „pedestrian_area” (Tabela 4) dowodzi, że przyspieszenie to ma charakter stały i jest niezależne od wybranej wartości parametru Q . Dla pełnego przekroju badanych wartości ($Q=90, 60, 30$), model z jedną zmienną utrzymuje czas

wykonania na poziomie około 0,5 μ s, natomiast model z trzema zmiennymi oscyluje wokół 800-880 μ s, co w porównaniu do 2777-2972 μ s dla modelu odniesienia daje stałe, ponad trzykrotne przyspieszenie.

4.2. Całkowity czas kodowania obrazu

Redukcja czasu wyznaczania parametru λ musi być także odniesiona do całości czasu kodowania. Badanie redukcji czasu zostało przeprowadzone z uśrednieniem wyników dla większej liczby współczynników Q w obrębie trzech różnych zakresów (scenariuszy): Q_L dla $Q=\{20,25,30,35\}$, Q_M dla $Q=\{50,55,60,65\}$ oraz Q_H dla $Q=\{80,85,90,95\}$. W Tabelach 5 i 6 zaprezentowane zostały wartości przyspieszenia całego procesu kodowania z wykorzystaniem każdego z zaprezentowanych modeli dla dwóch wariantów kodowania Huffmana (standardowe i optymalne tablice kodów Huffmana) [10].

Tabela 5. Wartość przyspieszenia całego procesu kodowania dla kodowania wykorzystującego jedynie algorytm RDOQ

Scenariusz	Przyspieszenie względem odniesienia	
	Model 1 zmienna	Model 3 zmienne
Q_L	2,480x	1,772x
Q_M	2,005x	1,576x
Q_H	1,488x	1,305x

Tabela 6. Wartość przyspieszenia całego procesu kodowania dla kodowania wykorzystującego algorytm RDOQ oraz optymalizację tablic Huffmana (HTO)

Scenariusz	Przyspieszenie względem odniesienia	
	Model 1 zmienna	Model 3 zmienne
Q_L	2,309x	1,701x
Q_M	1,919x	1,531x
Q_H	1,471x	1,295x

Zmniejszenie narzutu obliczeniowego na wyznaczanie wartości parametru λ przekłada się na zauważalne przyspieszenie działania całego kodera. Na podstawie danych zaprezentowanych w Tabelach 5 i 6 (dla wariantów RDOQ oraz RDOQ+HTO) można zauważyć, że model z jedną zmienną generuje największe sumaryczne przyspieszenie, osiągając od około 1,47x w scenariuszu Q_H do blisko 2,48x w scenariuszu Q_L . Model z trzema zmiennymi oferuje mniejsze, ale wciąż znaczące zyski czasowe, na poziomie od 1,29x (Q_H) do 1,77x (Q_L). Różnica ta wynika wprost ze sposobu działania modelu z trzema zmiennymi, który do poprawnego wyznaczenia współczynników wymaga dodatkowego, wstępnego zakodowania obrazu w celu pobrania parametrów BPP i MSE. Jednak nawet przeprowadzenie tego wstępnego kodowania nie niweluje zysku wynikającego z wykorzystania modelu z trzema zmiennymi.

4.3. Analiza efektywności kompresji

Kolejnym aspektem wymagającym badania jest wpływ wykorzystania zaprezentowanych modeli na efektywność kodowania. Efektywność kodowania może się zmieniać, gdyż wartości parametru λ uzyskiwane z mo-

deli mogą w lepszy lub gorszy sposób odzwierciedlać rzeczywistą zależność (2) niż dzieje się to w koderze odniesienia. Wyniki zaprezentowane są w Tabelach 7 i 8, jako uśrednione wyniki dla takich samych scenariuszy, jak to zostało opisane w poprzednim podpunkcie. Raportowane są jakości liczone w mierze Bjøntegaard dRate [2] dla kanału luminancji (Y) oraz uśrednione dla luminancji i chrominancji (YCbCr).

Tabela 7. Efektywność kodowania określona jako delta Bjøntegaarda dla kodowania wykorzystującego jedynie algorytm RDOQ

Scenariusz	Model 1 zmienna		Model 3 zmiennie	
	dRate Y [%]	dRate YCbCr [%]	dRate Y [%]	dRate YCbCr [%]
Q _L	0,066	0,001	-0,254	-0,435
Q _M	0,011	0,124	-0,073	0,047
Q _H	-1,321	2,425	-0,144	-0,248

Tabela 8. Efektywność kodowania określona jako delta Bjøntegaarda dla kodowania wykorzystującego algorytm RDOQ oraz optymalizację tablic Huffmana (HTO)

Scenariusz	Model 1 zmienna		Model 3 zmiennie	
	dRate Y [%]	dRate YCbCr [%]	dRate Y [%]	dRate YCbCr [%]
Q _L	0,040	-0,065	-0,195	-0,509
Q _M	0,008	0,137	-0,050	0,068
Q _H	-1,350	2,441	-0,052	-0,152

Uproszczenie procesu wyznaczania mnożnika Lagrange'a niesie za sobą potencjalny koszt w postaci niewielkich odchyleń w końcowej efektywności kompresji. Wyniki zestawione w Tabeli 7 (RDOQ) oraz 8 (RDOQ+HTO) wykazują, że model z trzema zmiennymi charakteryzuje się efektywnością znacznie bliższą modelowi odniesienia, osiągając niższe lub ujemne wartości dRate. Oznacza to poprawę bądź utrzymanie wysokiej efektywności. W szczególności dla kanałów YCbCr w modelu z trzema zmiennymi, wahania wskaźnika dRate wynoszą od -0,509% do +0,068%. Model z jedną zmienną generuje gorsze wyniki wydajności kompresji w scenariuszach wysokiej jakości (Q_H), gdzie strata dla kanałów YCbCr sięga około 2,4%. Dla niższych jakości strata wynikająca ze stosowania modelu z jedną zmienną jest znacznie mniejsza, a w skrajnym przypadku (Q_L) daje nawet bardzo niewielką poprawę względem modelu odniesienia.

5. PODSUMOWANIE

Zastosowanie modeli zaprezentowanych w artykule pozwala na wyraźne przyspieszenie kodowania obrazów do formatu JPEG z wykorzystaniem optymalizacji RDOQ. W części rozważanych scenariuszy poza przyspieszeniem kodowania uzyskuje się również poprawę efektywności kodowania obrazu. Wykorzystanie modelu z jedną zmienną jest rekomendowane w aplikacjach priorytetyzujących maksymalną szybkość działania, podczas

gdymodel z trzema zmiennymi stanowi optymalny kompromis między znacznym przyspieszeniem koderaprzy zachowaniu założonej prędkości bitowej.

PODZIĘKOWANIA

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

LITERATURA

- [1] Bampis C. G., Li Z., Moorthy A. K., Katsavounidis I., Aaron A., Bovik A. C. 2016. „LIVE Netflix Video Quality of Experience Database”.
- [2] Bjøntegaard G. 2008 „Improvements of the BD-PSNR model,” *Technical Report, VCEG-AI11*.
- [3] Bronsztejn I.N., Siemiendiajew K.A., Musiol G., Muhlig H. 2007. „Nowoczesne kompendium matematyki, Wydawnictwo Naukowe PWN.
- [4] Chlubna T., Zemčík P. 2025. „Comparative survey of image compression methods across different pixel formats and bit depths”. *Signal, Image and Video Processing*, 19.
- [5] Dornauer B., Felderer M. 2023. „Web Image Formats: Assessment of Their Real-World-Usage and Performance Across Popular Web Browsers”. *International Conference on Product-Focused Software Process Improvement*.
- [6] Fischler M. A., Bolles, R. C. 1981. „Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”. *Communications of the ACM*. 24 (6) : 381–395.
- [7] Grajek T., Stankowski J. 2022. „Rate–distortion optimized quantization in Motion JPEG,” *International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*.
- [8] ISO/IEC 10918-1. 1992. „Information Technology — Digital Compression and Coding of Continuous-Tone Still Images — Requirements and Guidelines”, ITU-T Rec. T.81.
- [9] Savitzky A., Golay M. J. E. 1964. „Smoothing and Differentiation of Data by Simplified Least Squares Procedures”. *Analytical Chemistry*. str. 1627-1639.
- [10] Stankowski J., Fojut A., Karwowski D., Grajek T. 2025. „Łączne wykorzystanie techniki RDOQ i metody optymalizacji kodów Huffmana w koderze JPEG”. *Przegląd Telekomunikacyjny - Wiadomości Telekomunikacyjne*. 4 : 463-466.
- [11] Stankowski J., Klimaszewski K., Grajek T., Grzelka A. 2026. „Parallel JPEG encoder with Rate Distortion Optimized Quantization (RDOQ) (and other optimizations)”. <https://github.com/jstankowski/PJRD>.
- [12] Stankowski J., Śliwiński A., Grzelka A., Grajek T. 2024. „Analiza możliwości zrównoleglenia obliczeń w koderze i dekodecie JPEG”. *Przegląd Telekomunikacyjny - Wiadomości Telekomunikacyjne*, 4 : 438-441.
- [13] W3Techs, 2026. „Usage statistics of image file formats for websites”. https://w3techs.com/technologies/overview/image_format.