

局所的な映像特徴を考慮した R-D 最適化の性能改善に関する一検討

吉野 知伸 内藤 整 小池 淳

株式会社 KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15

E-mail: {to-yoshino, sei, koike}@kddilabs.jp

あらまし IPTV や VOD をはじめ、HDTV コンテンツを IP ネットワーク上で配信するサービスの普及が期待されている。同サービスをコンシューマ向け広帯域回線として普及している DSL や FTTH において利用することを前提とすると、映像帯域は 10Mbps 以下に押さえられることが求められる。一方、MPEG-2 を用いて HDTV を符号化する場合、放送品質を満足するためには 20Mbps 程度のビットレートが必要であり、この条件を満たすためにはより高圧縮な符号化方式が求められる。一般に動画像の高圧縮符号化には MC+DCT 方式が適している。特に H.264/MPEG-4 AVC High Profile では HDTV に特化した高圧縮符号化ツールが規定されており、同方式は HDTV の圧縮効率に関して現在の標準化符号化方式の中で最も優れている。同方式で高い符号化効率を実現するためには、符号化ツールの一つである符号化モードの適応選択や動き補償マクロブロックサイズ選択を正確に行うことが求められる。同選択の代表的な手法としては R-D 最適化法が知られ、近似的なラグランジュ乗数に基づく選択手法が用いられている。しかしながら、ラグランジュ乗数の実測値は入力素材によって異なるために近似的に同乗数を設定することは符号化効率の低下を招く恐れがある。以上の背景より、本稿では H.264/MPEG-4 AVC High Profile における RD 最適化法の性能改善により符号化効率が改善することの検証を目的とし、ラグランジュ乗数を実測により求め RD 最適化法に基づく符号化結果との比較を行う。また、ラグランジュ乗数の実測をフレーム単位で行い、ラグランジュ乗数の局所性を検証する。

キーワード H.264/MPEG-4 AVC, RD 最適化, HDTV

Study on Performance Improvement of R-D Optimization Technology Considering Local Image Feature

Tomonobu YOSHINO Sei NAITO and Atsushi KOIKE

KDDI R&D Laboratories Inc. 2-1-15 Ohara, Fujimino-shi, Saitama, 356-8502 Japan

E-mail: {to-yoshino, sei, koike}@kddilabs.jp

Abstract It is expected that HDTV broadcasting service over IP networks such as IPTV and VOD became popular. Assuming the use of high-speed internet service such as DSL and FTTH, the coding bit-rate should be limited up to 10Mbps. However, the required coding bit-rate to assure HDTV broadcasting quality by MPEG-2 technology is approximately 20Mbps. Therefore, more effective video coding technology is required to satisfy this requirement. Generally, MC + DCT hybrid coding technology is effective to achieve high coding efficiency. H.264/MPEG-4 AVC High Profile defines several coding functions to realize high coding efficiency especially for HDTV, and this standard now corresponds to the most effective video coding scheme for HDTV among international video coding standard. In this standard, precise mode decision on Frame/Field coding mode selection, which is one of the coding functions to realize high coding efficiency, and MC (Motion Compensated interframe prediction) macroblock size selection is required to achieve high coding efficiency. In the conventional R-D optimization scheme, the cost function is defined by approximating the Lagrange coefficient. However, the Lagrange coefficient should be calculated dependent on image feature, and this approximate coefficient might cause performance degradation. From this perspective, we clarified performance improvement of R-D optimization scheme for H.264/MPEG-4 AVC High Profile coding. The difference in Lagrange coefficient between the actual measurement and the approximating scheme was confirmed. Moreover, we calculated the actual Lagrange coefficient by frame, and locality of the Lagrange coefficient was confirmed.

Keyword H.264/MPEG-4 AVC, R-D Optimization, HDTV

1. はじめに

IPTV や VOD をはじめ、HDTV コンテンツを IP ネットワーク上で配信するサービスの普及が期待されている。同サービスをコンシューマ向け広帯域回線として普及している DSL や FTTH において利用することを前提とすると、映像帯域は 10Mbps 以下に抑えられることが求められる。しかし、MPEG-2 を用いて放送品質を満足しつつ HDTV を符号化するためには 20Mbps 程度のビットレートが必要となるため、前述のサービスを促進するためにはより高圧縮な符号化方式が求められる。

一般に動画像の高圧縮符号化には MC+DCT 方式が適しており、H.264/MPEG-4 AVC[1](以下、H.264)は、可変ブロックサイズ動き補償に代表される高圧縮符号化ツールにより、現在の標準化符号化方式の中で最も高い符号化効率を実現する。特に High Profile では HDTV に特化した高圧縮符号化ツールとして符号化モードの適応選択が規定されており、HDTV の符号化効率に関して最も優れている。

同方式における符号化ツールの規定はデコーダに関するものであり、同ツールにおける符号化モードや可変ブロックサイズの選択方法はエンコーダ側に委ねられている。よって高い符号化効率を得るためには、エンコーダにおいて同ツールに関する選択を正確に行うことが求められる。本稿では、性能向上により大きなゲインが期待される同ツールにおける選択手法に着目する。

同選択における代表的な手法として RD 最適化法[2]があり、近似的なラグランジュ乗数を用いたコスト関数に基づく選択手法が示されている。しかしながら、ラグランジュ乗数の実測値は入力素材によって異なるために近似的に同乗数を設定することは符号化効率の低下を招く恐れがある。

以上の背景より、本稿では H.264 High Profile において、入力素材に対して最適なラグランジュ乗数を得ることにより符号化効率が改善することの検証を目的としている。文献[2]に準じて実測のラグランジュ乗数を算出して文献[2]におけるラグランジュ乗数と比較すると共に、フレーム単位で実測のラグランジュ乗数を求め、同乗数の局所性に関する検証を行ったので報告する。

2. シーケンス単位のラグランジュ乗数の実測

2.1. 実測によるラグランジュ乗数の導出

式(1)にラグランジュ乗数を用いたコスト関数[2]を示す。ただし、 J はコスト値、 D は符号化による誤差信号、 R は発生符号量、 λ は量子化ステップサイズを

表している。また、 λ はラグランジュ乗数であり、 λ の関数である。式(1)は、可変ブロックサイズの選択や符号化モードの適応選択において、ある λ に対して最小の J が得られる候補を選択することにより、高い符号化効率を得られることを表している。

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

文献[2]における λ の算出は、H.263[3] 符号化方式において複数素材について実測の R, D に対し、適切な λ を近似的に求めることにより行われている。

しかしながら、 λ の実測値は入力素材によって異なる。実測値の λ を設定することによる符号化効率の改善効果を検証するために、文献[2]に準じて H.264 High Profile において、実測により λ を λ の近似関数(以下、 λ - λ 曲線)として求める。実測 λ - λ 曲線を導出する手順を以下に示す。

Step 1. N 個の任意の正数 $\lambda = \lambda_n$ ($n=1,2,\dots,N$) を選択する。

Step 2. λ_n に対して任意の量子化ステップサイズを用いて符号化を行い、ピクチャタイプごとに式(1)の値が最小となる量子化ステップサイズ λ_n を求める。

Step 3. $n=1,2,\dots,N$ について Step.2 を行い、ピクチャタイプごとに λ_n, λ_n ($n=1,2,\dots,N$) を得る。

Step 4. Step.3 の結果より、ピクチャタイプごとに λ - λ の近似関数を求める。

上記の手順に従い、 $\lambda=25,50,100,180,400$ に対して λ - λ 曲線を 2 次回帰曲線として求めた。エンコーダとして、JM(Joint Model)10.1[4]を用いた。また、符号化対象として、ITE HDTV 標準動画像の 10 素材(シーケンス番号 7, 12, 16, 19, 20, 23, 25, 30, 43, 46)のそれぞれ 0 フレーム目から 29 フレーム目を用いた。符号化条件を表 1 に示す。得られた λ - λ 曲線および、JM における λ - λ 曲線を図 1 から図 3 に示す。ただし JM では文献[2]の手法に基づく λ - λ 曲線が用いられている。図 1 から図 3 はそれぞれ I ピクチャ、P ピクチャ、B ピクチャの結果を示している。また、一連の図は 2 素材(シーケンス番号 16, 46)の結果を示している。結果より、I ピクチャについては実測の λ - λ 曲線と JM における λ - λ 曲線がほぼ一致している一方で、P ピクチャおよび B ピクチャについては実測に対して JM の λ - λ 曲線が乖離していることが分かる。なお、図 1 から図 3 に示した以外の 8 素材における λ - λ 曲線についても、同様の傾向が見られた。

表 1 符号化条件

設定項目	設定値
プロファイル	High Profile
レベル	4.0
R-D最適化	あり
DCT変換サイズ	4×4 および 8×8
符号化方式	CABAC
アダマール変換	なし
動き予測	UMHexagon Search
符号化モード	MB単位 Frame/Field符号化モード適応選択
GOP構成	IBBPBBPBBPBBPBB
参照ピクチャ	IおよびPピクチャ
参照ピクチャ数	2
動きベクトル探索範囲	±64 × ±64

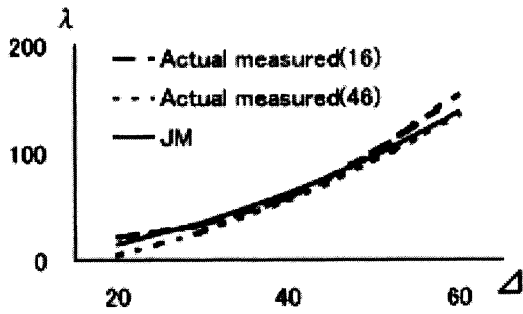


図 1 λ-Δ曲線(Iピクチャ)

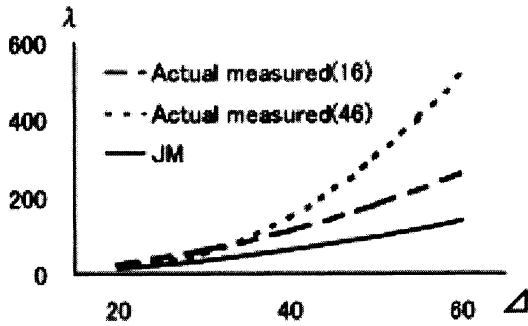


図 2 λ-Δ曲線(Pピクチャ)

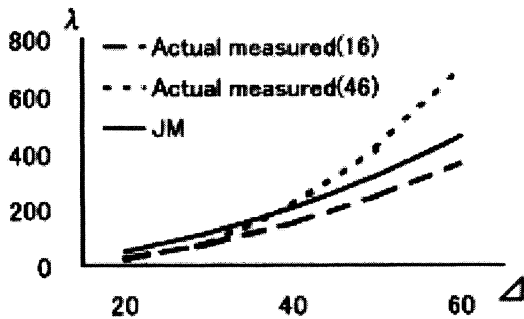


図 3 λ-Δ曲線(Bピクチャ)

2.2. 符号化実験

2.1 で用いた素材に対し、実測で得られた λ - Δ 曲線を用いて符号化を行った。符号化条件は 2.1 と同様である。符号化結果および JM の符号化結果として、シーケンス番号 46 に対する R-D 特性を図 4 に示す。なお、他の素材については、10Mbps における SNR[dB] を表 2 に示す。結果から、 λ - Δ 曲線を実測で求めることによるゲインは入力素材によって異なり、最大で 0.1dB 程度の符号化ゲインが得られていることが分かる。

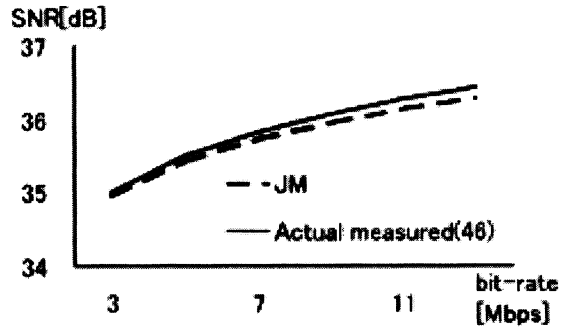


図 4 R-D 特性

表 2 符号化実験結果

シーケンス番号	実測 λ - Δ 曲線 SNR [dB]	JM SNR [dB]
7	33.23	33.21
12	33.53	33.52
16	28.75	28.67
19	28.57	28.54
20	28.29	28.27
23	23.73	23.71
25	43.45	43.28
30	35.94	35.88
43	32.87	32.82
46	36.16	36.06

3. フレーム単位でのラグランジュ乗数の実測

ラグランジュ乗数は素材に依存して変化することが知られ、図 1 から図 3 からわかる。2 章では文献 [2] に準じて、各素材につき素材全体に対してピクチャタイプごとに λ - Δ 曲線(以下、シーケンス λ - Δ 曲線)を導出したが、本章ではより細かい単位で λ - Δ 曲線を得て、ラグランジュ乗数の局所性について検証を行う。

3.1. 実測によるラグランジュ乗数の導出

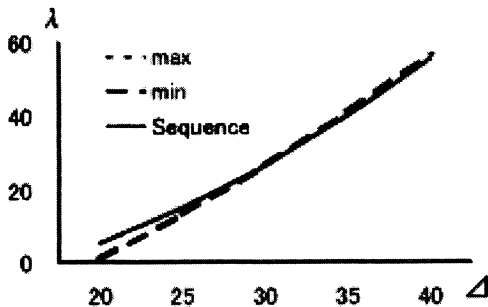
2 章に示した H.264 High Profile における実測 λ - Δ 曲線導出手順に準じて、フレームごとに λ - Δ 曲線を得る。以下に手順を示す。

- Step 1. N 個の任意の正数 $\lambda = \lambda_n$ ($n=1,2,\dots,N$) を選択する。
- Step 2. λ_n に対して任意の量子化ステップサイズを用いて符号化を行い、フレームごとに式(1)の値が最小となる量子化ステップサイズ Δn を求める。
- Step 3. $n=1,2,\dots,N$ について Step.2 を行い、フレームごとに λ_n , Δn ($n=1,2,\dots,N$) を得る。
- Step 4. Step.3 の結果より、フレームごとに λ - Δ の近似関数を求める。

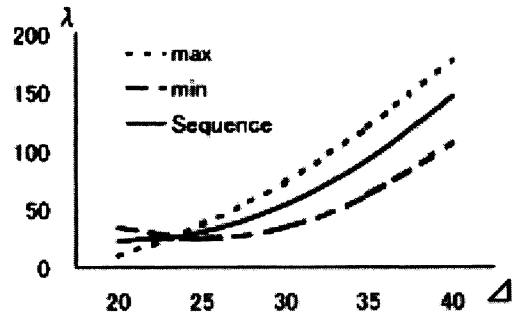
上記の手順に従い、 $\lambda=25,50,100,180,400$ に対して λ - Δ 曲線を 2 次回帰曲線として求めた。符号化条件および符号化対象は 2 章と同様である。

フレーム単位の λ - Δ 曲線および、シーケンス λ - Δ 曲線を図 5, 図 6 に示す。図 5, 図 6 はそれぞれシーケンス番号 46, シーケンス番号 23 における結果を示し、各図において I ピクチャ, P ピクチャ, B ピクチャの結果を示している。なお、図中の max, min はそれぞれ λ - Δ 曲線が最も大きいフレームおよび最も小さいフレームにおける λ - Δ 曲線を表している。また、シーケンス λ - Δ 曲線を Sequence として表している。結果より、素材によってフレームごとの λ - Δ 曲線の変化の大きさが異なり、図 5 では P ピクチャ, B ピクチャの λ - Δ 曲線がシーケンス λ - Δ 曲線から乖離していることが分かる。一方、図 6 では全てのピクチャタイプにおいて、フレーム単位とシーケンス λ - Δ 曲線がほぼ一致しており、I ピクチャ, P ピクチャにおいては全てのフレームがシーケンス λ - Δ 曲線に一致している。

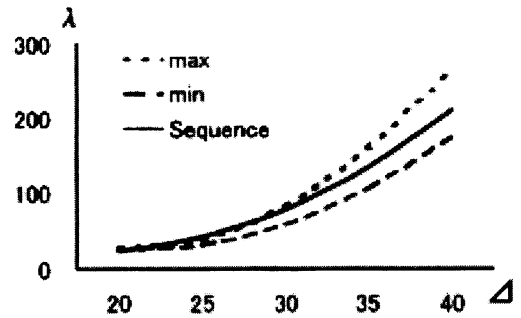
なお、図 5, 図 6 に示した以外の 8 素材における λ - Δ 曲線については、シーケンス番号 7, 30, 43 などでは図 5 と同様の傾向が得られ、シーケンス番号 25 などでは図 6 と同様の傾向が見られた。



(a) I ピクチャ

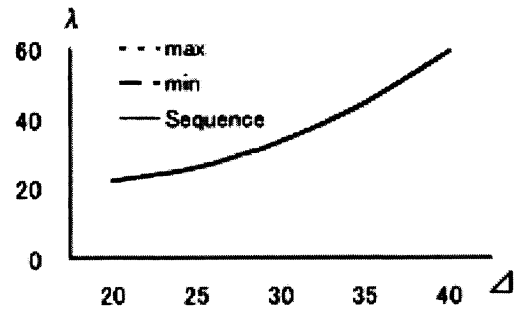


(b) P ピクチャ

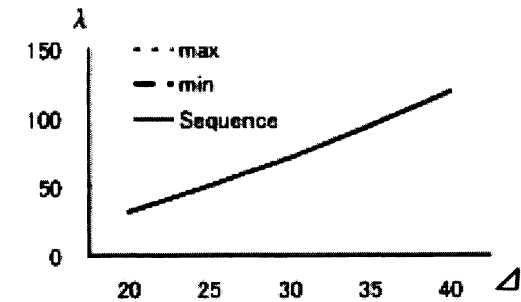


(c) B ピクチャ

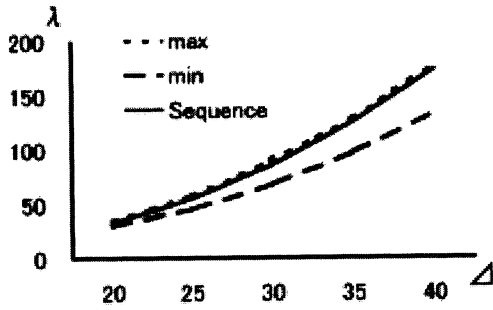
図 5 フレーム単位の λ - Δ 曲線(シーケンス番号 46)



(a) I ピクチャ



(b) P ピクチャ



(c) B ピクチャ

図 6 フレーム単位の λ - Δ 曲線(シーケンス番号 23)

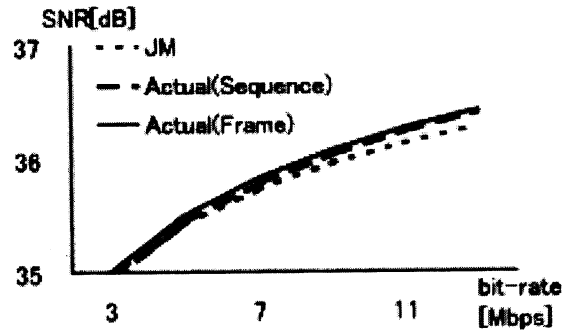


図 7 R-D 特性

3.2. 符号化実験

3.1 で得られた λ - Δ 曲線を用い、2 章で用いた素材に対して符号化実験を行った。符号化条件は 2 章と同様である。符号化結果として、シーケンス番号 46 の RD 特性を図 7 に示す。なお、図中の Sequence はシーケンス λ - Δ 曲線を用いて符号化した結果であり、Frame は 3.1 で求めたフレームごとの λ - Δ 曲線を用いて符号化した結果を表している。図 7 に示した以外の素材については、10Mbps における JM に対する SNR のゲインおよびシーケンス λ - Δ 曲線を用いて符号化した結果に対する SNR のゲインを表 3 に示す。結果より、 λ - Δ 曲線をフレーム単位とすることにより、シーケンス番号 46 のようにフレームごとに λ - Δ 曲線が大きく変化する素材においては、シーケンス λ - Δ 曲線の結果に対して 0.1dB 程度の符号化効率が改善している。一方、シーケンス番号 23 のようにフレームごとの λ - Δ 曲線の変化が小さい素材においては、シーケンス λ - Δ 曲線の結果とほぼ同等にとどまっている。

ここで、素材の映像特徴とラグランジュ乗数の局所性について考察する。シーケンス番号 25 のようにほぼ静止している素材や、シーケンス番号 23 のように単調なカメラワークかつフレーム間で絵柄が似通った素材においては、フレーム間で映像特徴の局所性が低い。よってラグランジュ乗数はフレームごとの局所性が低く、フレームごとの λ - Δ 曲線の変化が小さいと考えられる。

一方、シーケンス番号 16 のように特定の物体を追従するカメラワークの素材や、シーケンス番号 30 のように固定カメラに対して複数の人物が複雑に動く素材においては、フレーム間で映像特徴に局所性が高い。よってラグランジュ乗数はフレームごとの局所性が高く、フレームごとの λ - Δ 曲線の変化が大きいと考えられる。

表 3 符号化実験結果

シーケンス番号	JM に対するゲイン [dB]	2 章に対するゲイン [dB]
7	0.10	0.08
12	0.02	0.01
16	0.10	0.02
19	0.06	0.03
20	0.05	0.03
23	0.02	0.00
25	0.18	0.01
30	0.08	0.02
43	0.15	0.10
46	0.13	0.03

4. ラグランジュ乗数の推測に関する考察

2 章および 3 章では、ラグランジュ乗数の実測値を得るために符号化対象のフレームを複数回にわたって符号化を行うため、一連の処理には多大な計算コストが強いられる。

本章では、ラグランジュ乗数の局所性を踏まえて、過去の符号化結果から当該フレームにおけるラグランジュ乗数を推測することに関する考察を行う。

4.1. 誤差信号と発生符号量の比とラグランジュ乗数

文献[2]のコスト関数におけるラグランジュ乗数は、符号化において発生する符号量に対して想定される誤差信号の比を表している。実際の符号化における誤差信号と符号量の比(以下、D/R 比)は、符号化結果の符号量と SNR から算出することができる。

ここで、実際の符号化における RD 特性は、同一符号化条件であってもモード選択手法によって異なり、コスト関数を用いる手法においては適用するラグランジュ乗数によって異なるものの、実測のラグランジュ乗数と同様に符号化対象とする素材に大きく依存する。したがって、D/R 比はラグランジュ乗数の実測値と相関性を持つと考えられる。図 8 は、シーケンス 20 につ

いて、QP=27, 28, 29, 30, 31 に対して JM を用いて符号化した結果の D/R 比と、シーケンス λ - Δ 曲線において該当する量子化ステップサイズに対する λ の組をプロットした図である。なお、符号化条件は 2 章と同様である。図より、D/R 比と実測のラグランジュ乗数と同様の値を示しており、両者に相関性があると言える。すなわち、同一符号化条件の下で過去の符号化結果における D/R 比を用いることにより、シーケンス λ - Δ 曲線から得られる λ に相当するラグランジュ乗数を予測することができると考えられる。

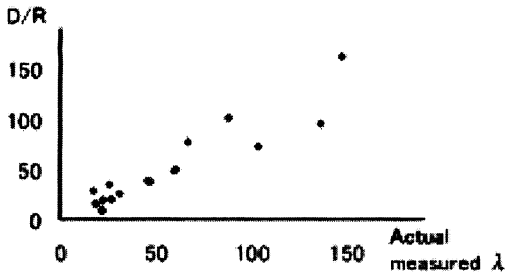


図 8 D/R 比と実測のラグランジュ乗数

4.2. 該当フレームにおけるラグランジュ乗数の推測

一定のモード選択手法の下で符号化を行った結果について、4.1 よりラグランジュ乗数の実測値と D/R 比は相関性を持つことから、各フレームの D/R 比は当該フレームの映像特徴に依存し、当該フレームにおける実測のラグランジュ乗数と相関を持つと考えられる。

図 9 に、シーケンス番号 16 およびシーケンス番号 23 について、B ピクチャにおける誤差信号と発生符号量の比をフレームごとに示す。符号化条件は 2 章と同様であり、QP=29 としてエンコーダは JM を用いた。結果より、シーケンス番号 23 の 13 フレーム目において D/R 比の大きな変化が見られるものの、それ以外においては D/R 比は時間的に相関性を持って推移している。

以上より、D/R 比はラグランジュ乗数の実測値と相関性を持ち、時間的な相関性も高いことから、直前の同ピクチャタイプにおける D/R 比を用いて、当該フレームにおける実測値相当のラグランジュ乗数を推測することが可能であると考えられる。

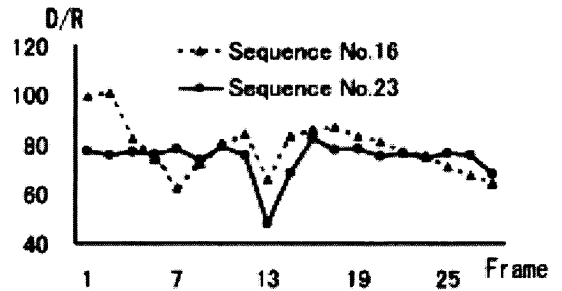


図 9 D/R 比

5. まとめ

本稿では、H.264 High Profile における RD 最適化法[2]の性能改善に関する検証を目的とし、ラグランジュ乗数を符号化対象の素材全体に対して実測で求め、文献[2]に基づく近似値と実測値が乖離していることを確認した。実測のラグランジュ乗数を用いて符号化を行うことにより、入力素材によってゲインが異なるものの、最大で 0.1dB 程度の符号化効率が改善することを確認した。

更に、ラグランジュ乗数の局所性を検証するために、同乗数をフレーム単位で実測にて求め、シーケンス λ - Δ 曲線と乖離する場面があることを確認した。同傾向の素材においては、フレーム単位の実測のラグランジュ乗数を用いて符号化を行うことにより、文献[2]に対して 0.15dB 程度の符号化効率が改善することを確認した。

また、符号化結果における D/R 比とラグランジュ乗数の実測値の相関性を検証し、同性質を用いたラグランジュ乗数の推測が可能であることを確認した。

なお、本研究は独立行政法人情報通信研究機構による委託研究「ソフトウェア符号化技術に関する研究開発」として実施したものである。

文 献

- [1] ITU-T Rec. H.264 | ISO/IEC 14496-10, "Advanced video coding for generic audio visual service", May 2003.
- [2] Gary J. Sullivan, Thomas Wiegand, "Rate-Distortion Optimization for Video Compression", IEEE Signal Processing Magazine, pp.74-90, Nov. 1998.
- [3] ITU-T (formerly CCITT), "Video coding for low bitrate communication", ITU-T Recommendation H.263; version 1, Nov. 1995; version 2, Jan. 1998.
- [4] Karsten Suhling, H.264/AVC Software Coordination, <http://ip.hhi.de/index.htm>