

機関番号：13401

研究種目：基盤研究（C）（一般）

研究期間：2008～2010

課題番号：20500088

研究課題名（和文） 高精細動画像の圧縮符号化における時空間画質の最適制御に関する研究

研究課題名（英文） Research on the optimal control of spatio-temporal image quality in high-definition video coding

研究代表者

吉田 俊之 (YOSHIDA TOSHIYUKI)

福井大学・大学院工学研究科・教授

研究者番号：50240297

研究成果の概要（和文）： 与えられた伝送ビットレート下で動画像を帯域圧縮・符号化する際は、一般に時間-空間領域における画質の低下が生じる。本研究では、水平画素数 1000 を超え、毎秒のフレーム数 60 枚程度の高精細・高フレームレートの動画像を対象として、指定された伝送ビットレートで圧縮符号化した場合の主観画質を予測すると共に、その画質を最大化するための時間-空間領域におけるレート制御手法、さらには一定の伝送レートの下で画質の変動を最小化する画質平坦化法について検討し、高精細・高フレームレート動画像の圧縮符号化において有効な画質制御手法を確立した。

研究成果の概要（英文）： Video coding in a prescribed bit rate generally makes degradation on the image quality in the spatio-temporal domain. This research, focusing on high-definition (HD) videos whose horizontal size exceeds 1000 pixels and the frames per second is around 60, has investigated and developed an optimal rate control and video-quality equalization techniques based on an estimation of a subjective video quality under encoding in a prescribed bit rate.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	2,300,000	690,000	2,990,000
2009年度	300,000	90,000	390,000
2010年度	500,000	150,000	650,000
総計	3,100,000	930,000	4,030,000

研究分野：画像処理，信号処理

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：高精細動画像，圧縮符号化，画質，レート制御，MOS予測，画質平坦化

1. 研究開始当初の背景

一般に、動画像をネットワーク上で伝送する際は、利用可能な帯域に制限があるため、圧縮符号化が必要となり、また画質の劣化が不可避となる。従って、限られた帯域の中で画質劣化を最小にする、あるいは逆に画質を最大化する伝送を実現するためには、効率の良い圧縮符号化アルゴリズムを用いることはもちろんのこと、画質を最大化する最適制御が必要となる。動画像を圧縮符号化（以下、符号化という）する際の“困難さ”は、画像のアクティビティ、すなわち“単一フレーム内の高周波成分”や“フレーム間での動き量”に依存するため、非定常なアクティビティに対して適応的に画質を制御する必要がある。

この問題に対し、研究代表者らは、動画像のシーンを単位として主観画質のひとつである MOS (Mean Opinion Score) の予測値を最大化する手法を検討してきた（4章末の文献 [1-4]）。本手法では、符号化ビットレート (BR) を R [bits/s]、毎秒のフレームレート (FPS) を t [frames/s]、フレーム当たりの平均ビット量 (BPF) を s [bits/frame] としたとき、

$$R = s \cdot t \quad (1)$$

と書けることに着目し、図1のように、対象動画像を (s, t) の下で符号化した場合の MOS (0~4 の5段階スケール) の推定値 $EMOS(s, t)$ を求め、これを最大化する (s, t) の下で符号化する手法である。

文献 [1] では、 $EMOS(s, t)$ を

$$EMOS(s, t) \propto EMOS_s(s) EMOS_t(t) \quad (2)$$

のように、与えられた s に対して単一フレーム内の空間方向の MOS を推定する $EMOS_s(s)$ 、および与えられた t に対して動きの滑らかさに対する MOS を推定する時間方向の $EMOS_t(t)$ に分離可能なことを確認している。また、文献 [2], [3] では、 $EMOS_s(s)$ と $EMOS_t(t)$ を対象画像の特徴量から推定する手法を与えている。対象動画像のシーンを単位として $EMOS(s, t)$

が推定できれば、図1に示すように、与えられた BR R によって定まる双曲線上で $EMOS(s, t)$ が最大となる t (最適 FPS) および s の組合せの下で符号化すれば、限られた BR の中で画質を最大化する符号化が実現できる。文献 [4] では、このような時空間レート制御を実際に動画像符号化アルゴリズムに実装する詳細を論じている。

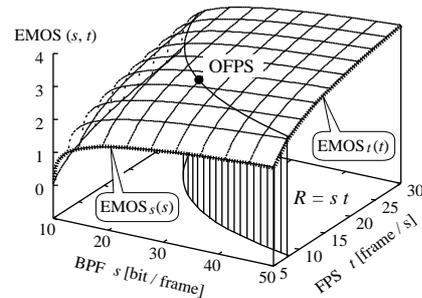


図1: $EMOS(s, t)$ の例と最適 FPS の決定

以上の時空間レート制御法は、これまで、サイズ 352×240 画素、フレームレート 30 [fps] 程度の、比較的小サイズの動画像に対して検討されてきたが、これを高精細・高フレームレート動画像 (HD 画像) へ適用することについては未だ検討されていない。また、複数シーンからなる動画像を定ビットレート (CBR) 下で符号化すると、画質はシーン間で大きく変動するが、一般に動画像全体の画質は最低画質のシーンに支配される傾向があるため、画質最低シーンの改善を図る必要がある。しかしながら、従来から画質評価に広く用いられている平均2乗誤差に基づく客観評価値では、基本的にシーン間の画質比較は不可能であるため、この問題は十分に検討されていない。

2. 研究の目的

以上の背景に基づき、本研究では、以下の3項目を研究の目的とした。

- (1) 1. で述べた時空間レート制御法を HD 画像に適用するための MOS の予測手法を確立する。
- (2) (1) に基づく時空間レート制御法を H.264 符号化器上で実現する。

- (3) シーン間の画質変動を抑える画質平坦化アルゴリズムを検討し、これを H.264 符号化器上で実現する。

3. 研究の方法

研究の目的 (1)–(3) に対応し、以下の方法に従って研究を進めた。

(1) HD 画像に対する MOS 予測法

①空間方向の予測式 $EMOS_s(s)$

研究代表者らが文献 [2] で提案した、SIF サイズの動画画像に対する MOS の予測法を HD 画像に拡張する。文献 [2] では、原画像–劣化画像間の PSNR [dB] 値と MOS の間には強い相関があることを利用し、空間方向の MOS 予測を s -PSNR の予測と PSNR–MOS の予測に分離している。前者は文献 [2] の方法で予測可能であるため、ここでは文献 [2] の手法に基づいて主観評価実験を行ない、後者の予測式を導出する。

②時間方向の予測式 $EMOS_t(t)$

同じく文献 [3] の手法を HD に拡張する。文献 [3] では、SIF サイズの動画画像に対し、フレームレートを低減する方法、および得られる動画画像の時間方向の MOS をその平均フレームレートから予測する方法が提案されている。ここでは、文献 [3] の方法を適用する際に必要となる、フレーム挿入の必要性を表す評価関数を HD 画像について導出し、これに基づいてフレームレートを低下させた画像の MOS を予測する予測式を導出する。

- (2) H.264 符号化器上での実現、および
(3) 画質平坦化の実現

(1) で得られた MOS の予測方法、およびそれに基づいて最適フレームレートをシーン単位で予測し、その下で符号化を行なう提案法を、H.264 符号化器の JM 参照ソフトウェア上に実装する。

また、複数シーンからなる動画画像を当該符号化器で CBR 符号化した場合の MOS

の変動を抑えるため、画質平坦化アルゴリズムを検討し、同じく JM 参照ソフトウェア上に実装する。また、複数の試験画像に適用し、その有効性を検証する。

4. 研究成果

本研究では、サイズ 1024[画素] × 768[画素]、フレームレート 60 [fps] の動画画像を HD 画像と位置づけ、これを利用して 2. の項目 (1)–(3) を実施した。

(1) HD 画像に対する MOS 予測法

①空間方向の予測式 $EMOS_s(s)$

11 種類の HD 試験画像を用いて主観評価実験を行ない、文献 [2] の手法を用いて空間方向の MOS 予測式

$$EMOS(x) = aT(x - T)^3 + bT(x - T) + c \quad (3)$$

$$T = d - e \ln(\varepsilon + 1) - f \ln(F + 1) \quad (4)$$

$$\begin{aligned} a &= 3.92 \times 10^{-4} \\ b &= 1.04 \times 10^{-2} \\ c &= 4.00 \\ d &= 5.38 \times 10^1 \\ e &= 4.35 \\ f &= 7.70 \times 10^{-2} \end{aligned}$$

を導出した。ここで、 x は対象画像–原画像間の PSNR 値 [dB]、 ε と F は対象画像から抽出される画像特徴量で、各々、エッジのフレーム間平均、フレーム間動きベクトル (MV) の最大値のフレーム間平均である [2]。11 種類の試験画像から順次 1 種類をオープンデータとして推定精度の検証を行なった結果、平均誤差 0.66、最大誤差 1.72 が得られ、ほぼ実用的な精度で空間方向の MOS が推定可能なことを確認した。

②時間方向の予測式 $EMOS_t(t)$

時間方法については、フレームを間引いてフレームレート t を低下させる手法、およびその手法で生成されたフレーム t のフレーム列に対して MOS を予測する手法の 2 つが必要となる。ここでは、文献 [3]

で提案されている，SD 画像に対する両手法を HD 画像に拡張した。

文献 [3] では，フレーム挿入の必要性を表す評価関数に基づいてフレーム間引きを実現しているため，まず文献 [3] の手法に基づいて 60 [fps] の HD 画像に対する評価関数 f を求めた。HD 画像の場合は， f は，フレームレート t [fps] と動領域の移動速度 m [pixels/frame] の他に，画面全体に占める動領域の割合 r に対する依存性が高いことが確認された。そこで，主観評価実験による結果を数量化し， m ， t ， r を用いて評価関数 f を

$$f(m, t, r) = a \left(1 - \frac{t}{60}\right)^{\frac{b}{r} + c} \cdot \tan^{-1} \left[\left\{ \frac{d}{t - e} + f \ln(g \cdot r) \cdot m \right\} \right] \quad (5)$$

$$\begin{aligned} a &= 7.70 \times 10^{-1} \\ b &= 3.80 \times 10^{-2} \\ c &= 1.34 \\ d &= 6.00 \times 10^{-3} \\ e &= 7.49 \\ f &= 1.59 \times 10^{-1} \\ g &= 1.14 \times 10^2 \end{aligned}$$

とモデル化した。式 (5) の評価関数を r ， t をパラメータとしてプロットした結果を図 2 に示す。

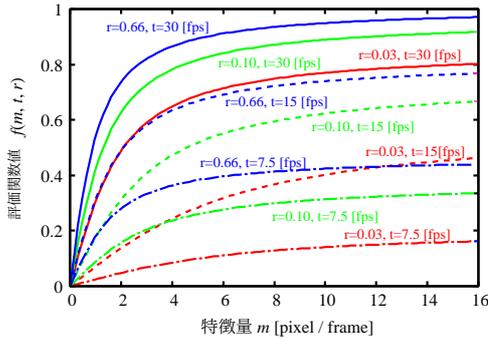


図 2：導出した評価関数 f

次に，文献 [3] のフレーム間引き手法を用いて複数の HD 試験動画のフレームレートを低下させ，その平均 FPS t_{avg} と主観評価実験で測定した時間方向の MOS の関係をモデル化し，時間方向の MOS 予測式として

$$EMOS(t_{avg}) = 4 \frac{\tan^{-1} b(t_{avg} - a)}{\tan^{-1} cb} \quad (6)$$

$$\begin{aligned} a &= 3.31 \times 10^{-1} \\ c &= 5.67 \times 10^1 \\ b &= -9.44 \times 10^{-2} \cdot r + 1.69 \times 10^{-1} \end{aligned}$$

を導いた。ただし，定数 b は動領域の割合 r に依存するため，文献 [3] の SD の場合とは異なり，EMOS も r に依存する。

(2) H.264 符号化器上での実現

(1) によって，HD 画像に対する時空間 MOS の予測が可能となったため，提案する時空間のレート制御法を実際の H.264 符号化器上に実装した。これは，マルチパスの符号化器で，最初のパスで対象動画から MOS 予測に必要な特徴量等を抽出して $EMOS(s, t)$ を構成し，次のパスで最適 FPS の予測を含む時空間レート制御を適用して符号化を行なう。

(3) 画質平坦化の実現

(2) で実現した符号化器を用いると，副産物として，符号化によって実現される MOS の予測値 (EMOS) が各シーン単位に得られる。対象動画を CBR 符号化すると，EMOS はシーン単位に大きく変動する。一般に，複数のシーンから成る動画の MOS は，最小 MOS を与えるシーンの MOS に支配される傾向にあるため，与えられた BR R の下で実現される画質を最大化するためには，最小 MOS の改善が必要となる。そこで，本研究では，大容量のバッファを用い，与えられた一定の平均伝送 BR R の下で符号化レートをシーン間で融通し合い，画質，すなわち EMOS の平坦化を図る手法を検討した。

まず，利用可能なバッファ量を B_{max} (単位：秒) とその初期占有量，画質平坦化の過程で各シーンが選択可能な BR (離散値) を規定する。いま，対象区間内には $i = 1, 2, \dots, n$ の n 個のシーンがあると仮定し， r_i を各シーンに与える BR として，

$$\max_{r_1, r_2, \dots, r_n} \min_{1 \leq i \leq n} EMOS_i(r_i) \quad (7)$$

$$\text{s.t.} \quad \sum_{i=1}^n r_i T_i \leq R \sum_{i=1}^n T_i \quad (8)$$

$$0 \leq B_i \leq B_{max} \quad (1 \leq i \leq N) \quad (9)$$

に基づくBR割り当てによって平坦化を行う。ここで、 T_i はシーン i の持続時間、 B_i はシーン i の符号化終了時点でのバッファ占有量 (単位: 秒) である。

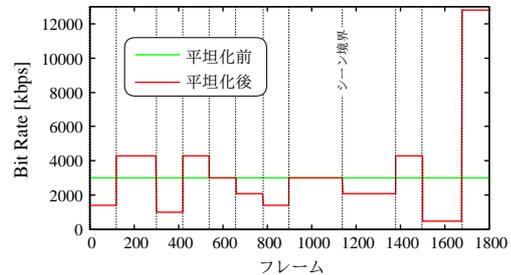
実際には、初期条件として各シーンに対して平均伝送BR R を与えた後、最も高いEMOSを与えるシーンのBRを1段階下げ、その過程で生じた余剰ビットを最小EMOSを有するシーンに割り当てる処理を、式(8)、(9)の条件下で繰り返すことで、式(7)を近似的に実行する。

本手法を、長さ1800フレーム(30秒)、各々12シーンからなるHD試験動画1, 2に適用し、画質平坦化の有効性を検証した。平均伝送BRを3000kbps, 各シーンが選択可能なBRを500~12800[kbps], バッファサイズを15秒(45Mbit)に設定し、各試験画像に対して画質平坦化を行なった結果を図3, 4に示す。ここで、両図の(a)はシーンを単位とするBRの推移、(b)はEMOSの変動を表しており、緑線が平坦化前、赤線が平坦化後である。式(7)–(9)を実行することによってシーン間でBRが融通される結果、シーンを単位にBR, EMOS共に変動し、最小EMOSが改善されて変動幅が低減されていることが確認できる。表1は平坦化の前後での最小EMOSを示しており、平坦化による最低画質の改善効果がわかる。

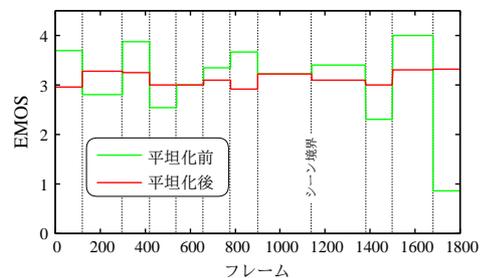
また、図5は、試験画像1および2に対し、バッファ量を変化させながら平坦化した場合の最大EMOS(赤線)、最小EMOS(青線)、および平均EMOS(緑線)の推移を表している。バッファ量を増加させると、最大EMOSは低下する一方で最小EMOSは徐々に改善し、その差は縮小する。これより、平坦化が進む過程が観察できる。また、バッファ量が画像依存の一定値に達すると平坦化の効果は飽和し、このバッファ量は試験画像1では12秒程度、同2では16秒程度である。試験画像1および2のシーン長の平均値は150フレームの2.5秒で、十分な平坦化を実現するためには、バッファ量は平均シーン長の数倍程度必要であることが結論できる。

表 1 : 平坦化前後での最小 EMOS

	平坦化前	平坦化後
試験動画画像 1	0.86	2.91
試験動画画像 2	1.21	2.51

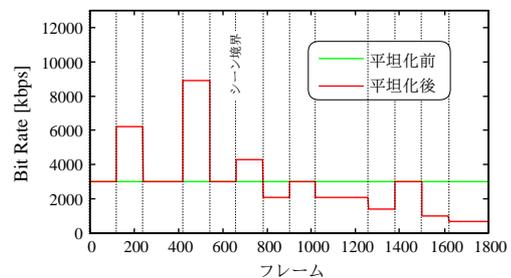


(a) 平坦化前後の BR の変動

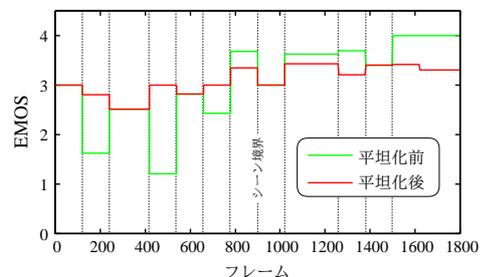


(b) 平坦化前後の EMOS の変動

図 3 : 試験動画画像 1 に対する平坦化の結果

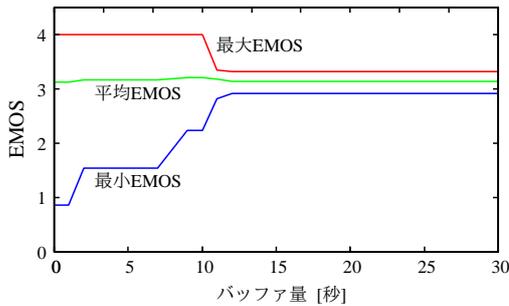


(a) 平坦化前後の BR の変動

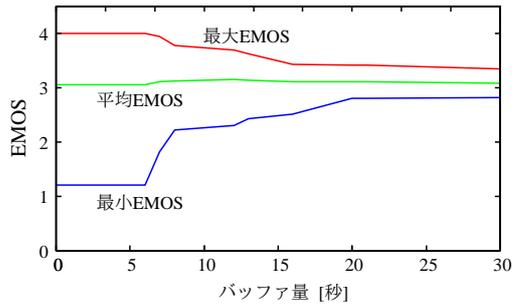


(b) 平坦化前後の EMOS の変動

図 4 : 試験動画画像 2 に対する平坦化の結果



(a) 試験動画像 1 に対する結果



(b) 試験動画像 2 に対する結果

図 5：バッファ量を変化させた場合の最大、最小、および平均 EMOS の推移

以上、本研究はほぼ当初目標の通り実施され、所期の成果が得られたと結論される。

【参考文献】

- [1] 稲積泰宏 他, “ビットレート制限下における動画像通信のための最適フレームレート推定”, 電子情報通信学会論文誌 (B), vol.85-B, no.7, pp.1130-1142 (2002 年 7 月)
- [2] 宮田高道 他, “SNR に基づく動画像の主観画質推定”, 電子情報通信学会論文誌 (A), vol.J88-A, no.11, pp.1292-1296 (2005 年 11 月)
- [3] 吉田俊之 他, “動画像に対する適応的フレーム間隔制御とその主観画質の予測”, 画像電子学会論文誌, vol.35, no.5, pp.497-508 (2006 年 9 月)
- [4] 吉田俊之 他, “主観評価値の最大化に基づく動画像の時空間レート制御とその符号化への応用”, 映像情報メディア学会論文誌, vol.62, no.11, pp.1785-1793 (2008 年 11 月)

5. 主な発表論文等

[雑誌論文] (計 1 件)

- ① 吉田俊之, “予測 MOS の平坦化に基づく符号化画質の改善法”, 電子情報通信学会論文誌 D, vol.J93D, no.9, pp.1634-1637 (2010 年 10 月)

[学会発表] (計 3 件)

- ① 山崎友也, 吉田俊之, “MOS 予測値の最大化に基づく高精細動画像のレート制御手法とその有効性について”, 映像情報メディア学会技術報告, vol.34, no.32, pp.47-52 (2010 年 7 月)
- ② 宮川大五郎, 吉田俊之, “予測 MOS の平坦化に基づく符号化画質の改善効果について”, 第 24 回画像符号化シンポジウム講演論文集, pp.35-36 (2009 年 10 月)
- ③ 田中伸一, 宮川大五郎, 吉田俊之, “バッファ占有量を考慮した MOS に基づく動画像画質の平坦化について”, 映像情報メディア学会技術報告, vol.32, no.35, pp.99-102 (2008 年 8 月)

6. 研究組織

- (1) 研究代表者
吉田 俊之 (YOSHIDA TOSHIYUKI)
福井大学・工学研究科・教授
研究者番号 ; 50240297
- (2) 研究分担者
なし
- (3) 連係研究者
なし