

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 7 日現在

機関番号：13401

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500112

研究課題名(和文) 動画像の階層型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 optimal control for spatio-temporal video quality in scalable video coding

研究代表者

吉田 俊之 (YOSHIDA, Toshiyuki)

福井大学・工学(系)研究科(研究院)・教授

研究者番号：50240297

交付決定額(研究期間全体)：(直接経費) 1,900,000円

研究成果の概要(和文)：SNRスケーラビリティ(各単一フレームの平均画質を改善)と時間スケーラビリティ(動きの滑らかさを改善)を有する動画像の階層(スケーラブル)符号化法において、主観的な画質評価指標であるMOS(Mean Opinion Score)の予測値(EMOS)に着目し、各階層でEMOSを最大化する画質制御手法を提案、確立した。また、シーン単位のEMOS変動を抑えるため、可変ビットレート符号化に基づく画質平坦化手法をスケーラブル符号化に導入し、その下で提案手法の有効性を総合的に検証した。

研究成果の概要(英文)：A video quality control technique for scalable video coding has been proposed and established in this research. The technique focuses on SNR-temporal scalable video coding, where SNR and temporal scalability gradually improves quality in terms of each video frame and motion smoothness, respectively, and maximizes an estimate for spatio-temporal mean opinion score (MOS) of each video scene in each scalable layer. A video quality equalization technique based on variable-bit-rate video coding has also been introduced in the proposed framework in order to reduce MOS fluctuation over scenes. Finally, total advantage of the proposed technique has been evaluated for several test sequences comprising multiple scenes.

研究分野：画像処理・信号処理

キーワード：動画像符号化 階層符号化 画質制御 レート制御 MOS

## 1. 研究開始当初の背景

動画像を送送／蓄積する際は、伝送可能な帯域あるいは蓄積可能なデータ量に圧縮するための圧縮符号化が行なわれる。動画像は、静止画像の時系列（各画像をフレーム、毎秒当たりのフレーム数をフレームレートと呼ぶ）であり、その総合的な画質は“各フレームの画質（空間画質）”と“フレームレートによって定まる動きの滑らかさ（時間画質）”によって定まる。

動画像を圧縮符号化する際は、目標となる帯域あるいはデータ量の中で、時間および空間の画質をバランスよく最大化する必要がある。その際の画質の評価量としては、数学的な扱いの容易さから、符号化-復号を経た復号画像と原画像との間の平均2乗誤差が主として用いられてきた（このような数学的誤差に基づく評価量を客観評価値と呼ぶ）が、客観評価値では画像間の画質を統一スケール上で一貫して評価できないなど、その限界が指摘されている。その対極的な評価尺度として Mean Opinion Score(MOS) が知られている。これは、複数の被験者に評価対象画像を提示して予め定められた評価スケール上の評価値を付けさせる“主観評価実験”を行ない、得られた評価値を被験者全員に渡って平均化した値である。動画像の画質を最終的に判断するのは人間であり、その意味で MOS に代表される主観評価値に基づく評価は極めて自然であるものの、MOS は主観評価実験を通じて得られる値であって、動画像を符号化する過程で MOS を最大化する制御を行なうことは不可能である。

以上の背景から、本研究代表者は従来より、対象動画像に対する MOS の予測値 (Estimated MOS: EMOS) を予測し、それを最大化する動画像符号化方式を提案してきた（4章末の文献[1-4]）。この方式は、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、

- FPS  $t$  は動きの滑らかさ（時間画質）
- BPF  $s$  は単一フレームの空間画質

を定め、さらに  $R$ ,  $s$ ,  $t$  の間には

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

の関係があることを利用する。具体的には図1に示すように、対象動画像のシーン（厳密は

時空間のアクティビティがほぼ一定な区間に分割したシーン）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。

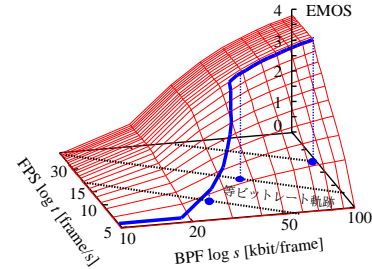


図1：EMOS( $s, t$ ) の例と最適 FPS の決定

文献 [1] では、EMOS( $s, t$ ) を

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

のように、与えられた  $s$ ,  $t$  に対して“空間方向の MOS を推定する  $EMOS_s(s)$ ”と“時間方向の MOS を推定する  $EMOS_t(t)$ ”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。

動画像各シーンの 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  $(s, t)$  の軌跡（図1は両軸対数のため、式(1)の下では直線群）上で EMOS( $s, t$ ) が最大となる  $t$ （これを最適 FPS と呼ぶ）、および対応する  $s$  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法を実際の動画像符号化器に適用する詳細を与えている。

一方、動画像を圧縮符号化する手法のひとつに、階層性を持たせたスケーラブル符号化 (SVC) 法が知られている。これは、複数の階層に分割されたビット列を順次復号することで階層に応じた画質の復号画像が得られる手法で、伝送ビットレートや演算能力が異なる heterogeneous な環境に対し、動画像を効率的に伝送する手法である。動画像の SVC には、

- 1) 時間スケーラビリティ：フレームレートに基づく時間画質を順次改善
- 2) SNR スケーラビリティ：2乗誤差 (SNR) に基づく空間画質を順次改善

### 3) 空間スケーラビリティ：画素数（解像度）を順次増加

の3種類が知られており、表示デバイスの解像度が一定の場合には1)と2)の組合せが可能であるが、各階層に与えられたBRの下で時間-SNRの画質を総合的に最大化する手法は検討されていない。

非階層的な符号化においては、EMOS最大化に基づく符号化法が確立されており、これを時間-SNR SVCに応用することで各階層の主観画質を最大化する最適レート制御が実現可能と考えられる。なお、本報告においては、過去の研究との連続性を考慮し、SNR スケーラビリティを特に空間 (SNR) スケーラビリティと呼ぶ。

## 2. 研究の目的

以上の背景に基づき、本研究では以下の3点を目的として研究を行なった。

(1) SVCにおいては時間スケーラビリティに伴うフレームレートによって符号化効率が変動するため、これを考慮した時空間のMOS予測式  $EMOS(s, t)$  を導出する

(2) 各階層でEMOSを最大化するスケーラビリティ選択手法/レート制御法を確立し、これをH.264/SVC階層符号化器上に実装する。さらに、シーン間でのMOS変動を抑えて動画全体画質向上を図る“画質平坦化法 [5]”を各階層に適用する。

(3) 得られた手法を実際の動画に適用し、提案手法の有効性を検証する。

## 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) 時空間 MOS 予測式 $EMOS(s, t)$ の導出

研究代表者らが文献 [2], [6] で提案、実現している空間MOS推定手法を用いる。この方法では、空間EMOSの予測にPSNRを介在させ、 $EMOS_s(s)$  を  $s$ -PSNR, PSNR-EMOSの各推定に分割して予測している。SVCの階層に依存した符号化効率変動は主として前者に影響するため、SVCに適用可能な  $s$ -PSNR 予測手法を中心に研究を行なった。一方の  $MOS_t(t)$  は基本的に文献 [3] の手法を利用した。

### (2) レート制御手法の確立と H.264/SVC 符号化器への実装/画質平坦化法の適用

#### ① レート制御手法の確立と実装

(1) で得られたEMOS予測手法を用いることで、各階層のビットレート（最下層からの累積ビットレート）に対応する時空間のMOS分布が予測されるため、これを各階層の時間あるいは空間 (SNR) スケーラビリティにマッピングする手法を確立し、スケーラビリティの選択手法/レート制御手法を実現する。

#### ② 画質平坦化法の適用

文献 [5] の画質平坦化法をSVC各階層に適用する。各階層毎に適切なサイズのバッファを持たせ、基本的に下位層から順次平坦化法を適用し、各層の平坦化を実現する。

### (3) 有効性の検証

提案手法を多くの試験動画に適用し、その有効性を検討する。提案法が有効に機能する画像を分析すると共に、平坦化あり/なしの結果を比較し、提案する時間-空間 (SNR) スケーラブル符号化を効果的に適用するための条件を検証する。

## 4. 研究成果

SVCにおいては、最下層をベース (B) 層、その上位各層をエンハンスメント (E) 層と呼ぶ。本報告ではB層を  $k = 0$ , E各層を  $k = 1, 2, \dots, N - 1$  層とする  $N$  層 SVC を考える。

時間スケーラビリティは、下位層のフレーム間に順次フレームを挿入することで実現されるため、挿入するフレーム数を各1枚と仮定し、原画像のFPSを  $t_0$ ,  $n$  を適当な整数とすると、各層のFPSは  $t_0 2^{-n}$  の値のみを取り得る。実用的には、時間スケーラビリティは最大で4層程度で最低FPSは7.5fpと仮定すると、提案手法においては例えば図2に示すフレーム間予測構造を用いることができる。

図2の  $k = 0, 1, 2$  層では、GOP(Group Of Picture) 内の (I, P, B) 各フレームの枚数は、各々(1, 2, 0), (1, 2, 3), (1, 2, 9) となっている。これを用いることで、以下の手順によって、 $s$ -PSNR の関係を階層毎に予測することができる：

1. 対象シーンを，数点の BR で SVC し，I/P/B 各フレーム毎の BR と得られる PSNR の関係をモデル式で近似する
2. 2. のモデル式を各層の (I, P, B) の枚数に従って加重和することで，任意の  $s$  に対する PSNR を予測する

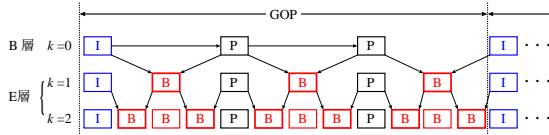


図 2：3 層 SVC におけるフレーム間予測構造の例

このような  $s$ -PSNR の関係予測，および文献 [2], [3], [6] の手法を用いることにより，SVC 各層における EMOS( $s, t$ ) の分布を図 1 のように予測することが可能となる．そこで提案法は，図 1 の最適 FPS 軌跡上で，第  $k$  層までの累積ビットレート  $R_k$  に対応する点 ( $s_k, t_k$ ) を求め， $s_k > s_{k-1}$  であれば空間 (SNR) 方向， $t_k > t_{k-1}$  であれば時間方向のスケラビリティを選択することを基礎とする．

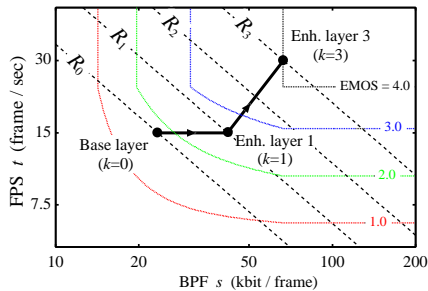


図 3：提案法におけるスケラビリティ選択

図 3 は，あるシーンに対し，EMOS( $s, t$ ) から得られる等 EMOS 線 (実線) を  $s$ - $t$  対数軸上にプロットした図で， $R_0 \sim R_3$  の各破線は  $N = 4$  層 SVC における各層の BR を表している．原画像の FPS を 30fps とすると，前述のように  $t_k$  が取り得る値は  $T = \{7.5, 15, 30\}$  のうちのひとつである．以上を基に，提案法では次の手順に従って各層のスケラビリティを選択しビット量を配分するレート制御を実現する．

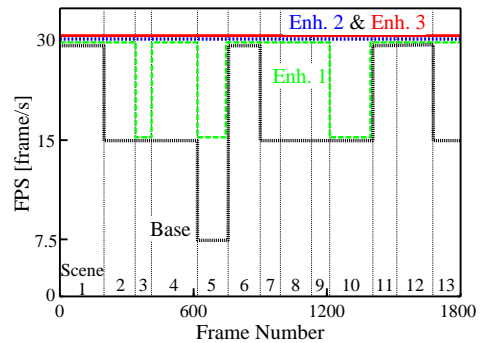
(1) B 層 ( $k = 0$ ) では，与えられた BR  $R_0$  に対して EMOS が最大となる  $t_0$  を  $T$  から選び， $(s, t) = (r_0/t_0, t_0)$  で符号化する．

(2) E 層 ( $1 \leq k < N - 1$ ) では，まず BR  $R_k$  に対して EMOS が最大となる  $t_k$ ，及び  $s_k = R_k/t_k$  を求める．ここで， $s_k < s_{k-1}$  となる場合は空間 (SNR) 方向への負のビット配分が必要で，これは不可能なため，

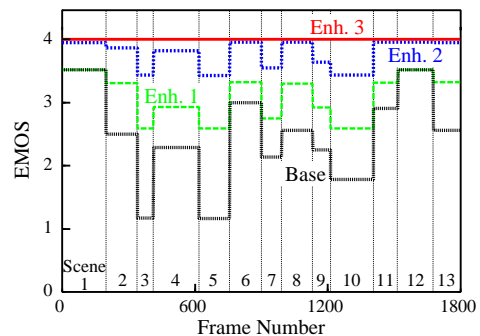
- 最上位の E 層 ( $k = N - 1$ ) では， $t_k \geq t_{k-1}$  かつ  $s_k \geq s_{k-1}$  を満たし，EMOS が最大となる  $t_k$  を選択し， $s_k = R_k/t_k$  で符号化する．
- 中間層 ( $1 \leq k < N - 1$ ) ではビット配分を保留し，当該層の BR を  $k + 1$  層に持ち越す．

図 3 の例では， $k = 1$  層では空間 (SNR) スケラビリティを選択して全ビット量を SNR 方向に配分， $k = 2$  では  $s_2 < s_1$  となるため全ビット量を持ち越し， $k = 3$  では時間と空間 (SNR) 両方向に配分している．

提案手法に対しては，さらに下位層から順に文献 [5] の画質平坦化法を適用することで，平均 BR 一定の下で各層の EMOS 変動を抑圧することができる．



(a) 各層各シーンに対する FPS の変動



(b) 各層各シーンに対する EMOS の変動

図 4：提案法の適用結果

以上の提案手法を H.264/SVC 符号化器上に

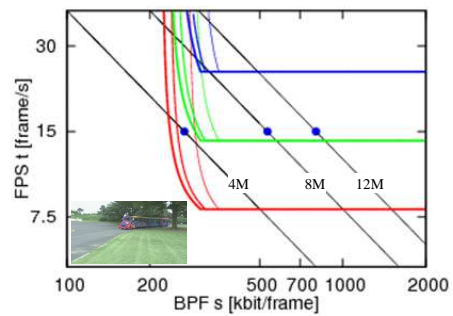
実装し、 $N = 4$  層 SVC によって有効性を検証した。実験には、13 シーンから成る SIF サイズ (352 × 240), 30fps, 1800 フレームの動画像を用い、各層の平均 BR として 350, 630, 1120, 2000[kbit/s] を与えた。また、画質平坦化法は、各層に 8.5 秒分のビット列を蓄積可能なバッファを用意し、その下で適用した。

図 4 は、各層各シーンに対する FPS の変動 (a), および EMOS の変動 (b) である。例えば、空間アクティビティが高いシーン 5 では、最大画質を得るために空間画質が優先され、FPS は B 層では低く抑えられている一方、上位層では時間スケラビリティによって徐々に改善されている。各シーンの EMOS についても、B 層から上位層へ向けて上昇すると共に、画質平坦化によってシーン間の変動が抑えられ、最上位層では一定 EMOS が実現されている。以上から、提案手法によって各層でシーンに適応したスケラビリティ選択/ビット配分が実現されていることが確認できる。

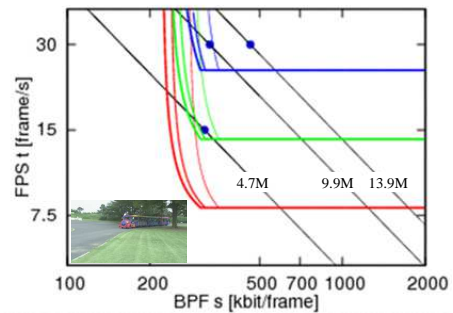
また、提案する時間-空間 (SNR) スケラブル符号化法が有効に機能する画像シーンを検証するため、10 シーンから成る HD サイズ (1920 × 1080), 30fps, 1500 フレームの動画像を用い、 $N = 3$  層で SVC 符号化した。ここでは、平均 BR 4M, 8M, 12Mbps を与え、平坦化ありとなしの場合を比較する。 $s-t$  平面において、各層の BR と等 MOS 線、および選択されるスケラビリティを図 5 に示す (等 MOS 線は、赤、緑、青が各々 EMOS=3.0, 2.0, 1.0 を表し、各 EMOS における 3 本の曲線は、各々左から 30fps, 15fps, 7.5fps に対する等 MOS 線である)。

図 5(a) に示す「シーン 2 平坦化なし」の場合は、層を上げても時間スケラビリティは選択されず、空間方向の画質のみが改善されている。これは、各層の BR が不十分なため、一方のスケラビリティのみが選択される例で、提案法はこの種の画像に対してはその有効性を十分に発揮できない。これに平坦化法を適用した結果 (b) では、各層の BR が上昇しており、その結果として時間と空間の両方向のスケラビリティが選択されるよう改善されている。一方、図 5(c), (d) では、平坦化なし/ありに関わらず、時間/空間 (SNR) 両方向のスケラビリティが適用されているため、提案法に取っ

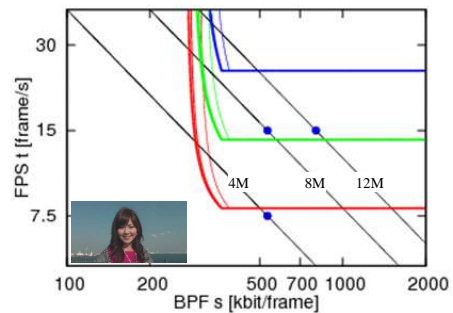
て有効な画像であると判断される。



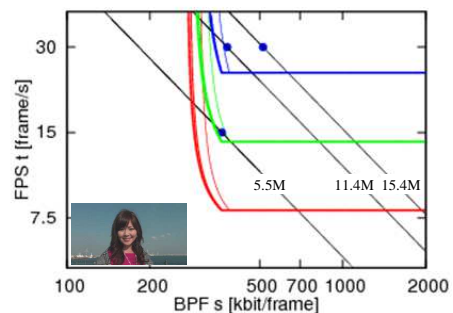
(a) シーン 2 の結果 (平坦化なし)



(b) シーン 2 の結果 (平坦化あり)



(c) シーン 2 の結果 (平坦化なし)



(d) シーン 2 の結果 (平坦化あり)

図 5 : スケラビリティ選択と提案法の有効性

以上から次の結論が導かれる。提案法は、EMOS を最大化する評価基準の下で時間/空間 (SNR) スケラビリティを最適に選択でき

る一方で、各層の平均 BR が不十分な場合は、一方のスケラビリティしか選択されず、提案法の有効性は減少する。一般に、シーンの時空間アクティビティはシーン間で大きく変動するため、各層を定ビットレート (CBR) 符号化する SVC は主観画質の観点からは十分ではなく、提案法を効果的に適用するためには、画質平坦化法を導入するなどして可変ビットレート (VBR) の下に適用する必要がある。

#### <引用文献>

- [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] 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.J93-D, no.9, pp.1634-1637 (2010 年 9 月)
- [6] 吉田俊之, “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”, 科学研究費補助金研究成果報告書, 基盤研究 (C) (一般), 課題番号 20500088 (2010 年 5 月)

#### 5. 主な発表論文等

[雑誌論文] (計 1 件)

- ① 吉田 俊之, “予測 MOS の最大化に基づく動画像階層符号化のための最適ビット

配分法”, 電子情報通信学会論文誌 D, 査読有, vol.J95-D, no.9, 2012, pp.1672-1674

[学会発表] (計 4 件)

- ① Shusuke Kondo, Yutaka Nakano, Toshiyuki Yoshida, “Effectiveness of scalable video coding based on a maximization of an estimated MOS”, International Workshop on Advanced Image Technology 2015, Jan. 13, 2015, Tainan (Taiwan)
- ② 近藤 秀祐, 仲野 豊, 吉田 俊之, “EMOS 最大化に基づく動画像スケラブル符号化法の有効性について”, 第 29 回画像符号化シンポジウム, 2014 年 11 月 13 日, ラフォーレ修善寺 (静岡県伊豆の国市)
- ③ Masamitsu Taue, Toshiyuki Yoshida, “Performance evaluation of a MOS estimation technique by means of SSIM”, International Workshop on Advanced Image Technology 2013, Jan. 7, 2013, Nagoya (Japan)
- ④ 田上 将充, 吉田 俊之, “画質評価における EMOS と SSIM の比較検討”, 第 27 回画像符号化シンポジウム, 2012 年 10 月 25 日, ニューウェルシティ湯河原 (静岡県熱海市)

#### 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