

映像シーンに適応したスケーラビリティ選択手法に関する検討

A study on scene adaptive scalability selection technique

高橋良知† 後藤崇行‡ 石川孝明‡ 渡辺裕‡ 富永英義†‡
 Yoshitomo TAKAHASHI Takayuki GOTO Takaaki ISHIKAWA Hiroshi WATANABE Hideyoshi TOMINAGA

† 早稲田大学理工学部 電子・情報通信学科

‡ 早稲田大学大学院 国際情報通信研究科

† Dept. of Elec.,Info. and Comm. Eng., Waseda Univ.

‡ Graduate School of GITS, Waseda Univ.

Abstract: The Scalable Video Coding (SVC), which is currently being developed as an extension tool of H.264/AVC, can provide temporal, SNR and spatial scalability. In this paper, we study the relation between the number of partitions of the temporal resolution and SNR in the highest frame rate. As a result, it is understood that the optimal number of partitions of the temporal resolution depend on the characteristics of the video sequence.

1 はじめに

現在 H.264/AVC の拡張ツールとして標準化が進められている SVC (Scalable Video Coding) では、空間・時間・SNR スケーラビリティを実現することで、ワンソースマルチユースを可能としている。しかし、一般的にスケーラブル符号化は非スケーラブル符号化に比べ符号化効率が低下することが知られている [1]。そこで、著者らは、動画像の特性に基づくスケーラブル符号化を行うことで、スケーラビリティを保ちつつ、符号化効率の低下を抑える検討を続けている。本稿では、このうち、時間スケーラビリティについて、時間解像度の分割数と最高フレームレートで復号した映像の SNR との関係、動画を特性と結びつけて検討を行う。

2 MCTF

2.1 MCTF の階層構造

SVC では、時間スケーラビリティを実現する手段として、MCTF (Motion Compensated Temporal Filtering) を用いる [2]。MCTF は図 1 のように、時間方向サブバンド分割を再帰的に実行することで、時間スケーラビリティを実現している。具体的には、GOP (Group of Pictures) 単位で、Prediction と Update の二つの操作を繰り返す。まず、Prediction により、残差信号を含んだハイパスフレームを生成する。その後、Update により、この残差信号を加算することによりローパスフレームを生成する。これらの Prediction と Update の操作には、Haar Wavelet と 5/3 Wavelet のいずれかが選択される。Haar Wavelet を用いた場合は単方向予測に相当し、5/3 Wavelet を用いた場合は双方向予測に相当する。デコーダでは、 $L_3, H_3, H_2, H_1, \dots$ の順にデコードすることで時間スケーラビリティを決定することができる。

2.2 GOP サイズと時間解像度階層数

図 1 は、GOP サイズが 8 の場合の MCTF の階層構造を示している。この場合、最高フレームレートから 3 階層の時間解像度を取り出すことが出来る。一般的に、 $2^N (N = 1, 2, 3, \dots)$ の GOP サイズでは、 N 階層の時間解像度が得られる。表 1 に、15(fps) の動画像に対する GOP サイズと取得可能なフレームレートの関係を示す。

Table 1: Relation between GOP size and extract frame rate

GOP size	Extract frame rate
2	15.0, 7.5
4	15.0, 7.5, 3.75
8	15.0, 7.5, 3.75, 1.875
16	15.0, 7.5, 3.75, 1.875, 0.9375

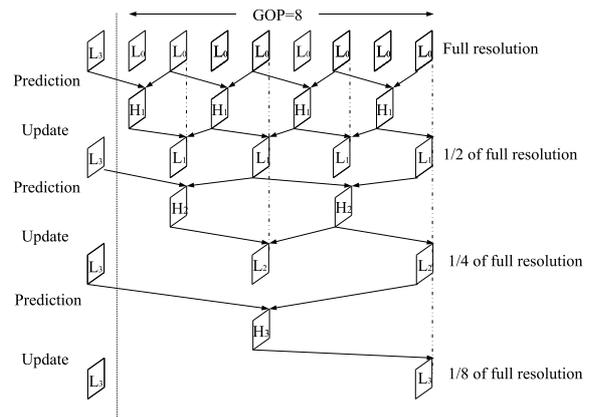


Figure 1: Structure of MCTF

3 予備実験

3.1 実験条件

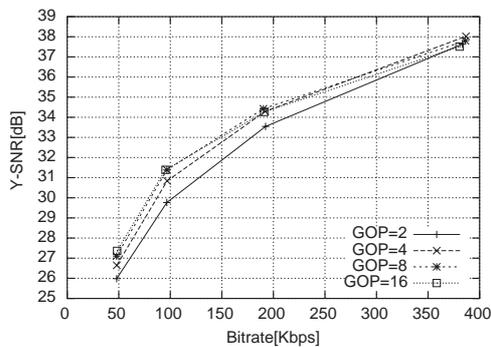
時間解像度の分割数と最高フレームレートでの SNR の関係を知るための予備実験を行った。予備実験は、SVC の参照ソフトウェア JSVM2.0 (Joint Scalable Video Model) [2] 上で行い、表 2 に示す実験条件に基づいた。なお、今回は MCTF の階層構造による影響を確認するために、Update の処理は省略した。

3.2 実験結果

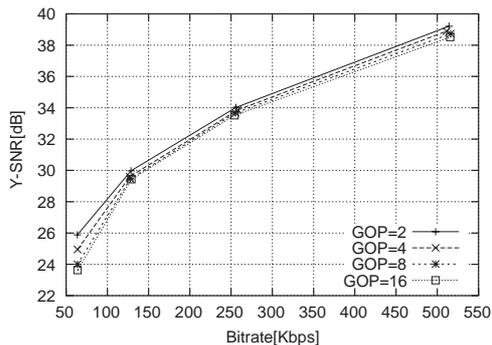
図 2 は、標準動画像 Mobile&Calendar, Football について、GOP サイズと最高フレームレートにおける SNR の比較を行ったものである。Mobile では GOP サイズが 8 の時に SNR が最大となっている。一方で、Football では、GOP サイズが 2 の時に SNR が最大となっている。このことから、特定の GOP サイズで最高フレームレートでの映像の SNR を最大にすることが分かる。さらに、その GOP サイズは動画像の特徴に依存して異なる。他の標準動画像については、Foreman, City は Mobile&Calendar と同様の特性を示した。また、Crew は Football と同様の特性を示した。

Table 2: Experimental Conditions

GOPsize	2, 4, 8, 16
Update Process	OFF
Number of Layers	1(QCIF,15fps)
Number of Frames	75
Intra Frame	Only First
Sequence	Mobile, Football, Crew, City, Foreman

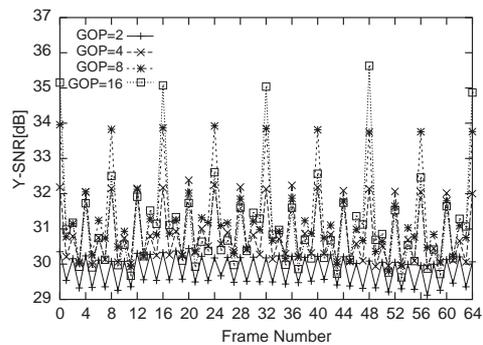


(a) Mobile

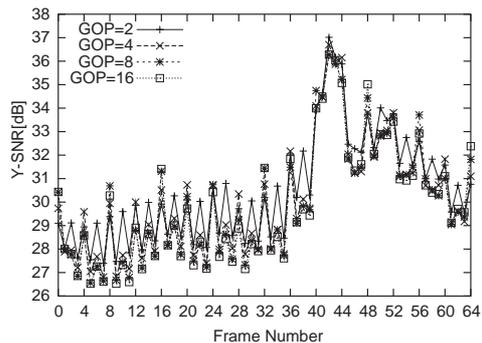


(b) Football

Figure 2: RD graph of different GOP size



(a) Mobile(96Kbps)



(b) Football(128Kbps)

Figure 3: Frame character of different GOP size

3.3 考察

実験結果に示されたような最大の SNR を示す GOP サイズが動画像により異なるのは、MCTF の階層構造のためである。具体的には、この階層構造に従い実行される動き補償と量子化の影響が GOP サイズにより異なる。

まず、動き補償では、図 1 のように、分割数が増えるほど、下位階層の時間解像度のローパスフレームは離れたローパスフレームを参照フレームとする構造となっている。このため、一般的に下位階層のハイパスフレームほど大きい残差を含んでいる。

また、JSVM では、下位階層の時間解像度から上位階層の時間解像度にかけて、量子化ステップサイズを次第に大きくしながら符号化を行う [2]。このため、階層の最下位に位置する L_0 フレームは最も小さい量子化ステップサイズで量子化される。一方で、階層の最上位に位置する H_1 フレームは最も大きい量子化ステップサイズで量子化される。したがって、フレームごとの SNR は図 3 に示すとおり、GOP サイズにより固有の変化を示す。

4 動画像の特徴と分割数

前節の考察より、エッジの少ない、激しい不規則的な運動を特徴とする動画像では時間解像度の分割数が少ないほど SNR が高く、エッジが多く含まれ、ゆるやかな規則的な運動を特徴とする動画像では時間解像度の分割数が多い

ほど SNR が高い、という特性があると考えられる。この特性を入力動画像の特徴と関係付けるため、TI(Temporal Infomation), SI(Spatial Infomation) を用いる [3]。TI とは、動きの激しさを表す特徴量である。また、SI は、エッジひずみやボケを測る空間方向の特徴量である。これらの特徴量を用いることで、入力動画像の性質を大まかに把握することが可能である。予備実験に用いた標準動画像の TI, SI を表 3 に示す。大きな GOP サイズで高い SNR を示す Football, Crew は、それぞれ、TI の最大値, SI の最小値を有している。しかし、これらの特徴量のみで最大の SNR を与える時間解像度の分割数を知ることは難しく、特徴量の重み付けや動きの規則性を捉えるパラメータなどの追加が必要である。

5 おわりに

本稿では、時間解像度の分割数と最高フレームレートでの映像の SNR の関係について検討を行った。その結果、最高フレームレートでの SNR を最大にする時間解像度の分割数は動画像の特性により異なることが確認できた。今後は、動画像に対して適応的な時間スケラビリティ及び SNR スケラビリティの選択手法に関する検討を行う。

参考文献

- [1] 木本, “MPEG におけるスケラブルビデオ符号化の標準化動向,” AVM-48, NO.10, pp.55-60, Mar.2004
- [2] J.Reichel, H.Schwarz, “Scalable Video Coding - Joint Scalable Video Model JSVM-2,” JVT-O201, Apri.2005
- [3] 岡本, 栗田, “入力映像の特徴量を考慮した映像品質客観評価法の検討,” 信学技報, CQ-52, MoMuC-39, P-61.66, 2003

Table 3: Value of SI and TI

	City	Crew	Mobile	Football	Foreman
SI	68	51	149	80	70
TI	238	348	562	1110	416