

## 動画像の品質主導型スケーラブル符号化 — 不均一フレーム間隔によるアプローチ —

宮田 高道 薄井 武順 吉田 俊之<sup>†</sup>

東京工業大学大学院理工学研究科集積システム専攻

〒 152-8552 目黒区大岡山 2-12-1

† 福井大学情報メディア工学科

〒 910-8507 福井市文京 3-9-1

E-mail : miyata@ip.ss.titech.ac.jp

あらまし SNR スケーラブル符号化の一種である FGS 方式はビット量制御に対して高い自由度を与える一方、従来型の時間スケーラブル符号化法はフレームレート選択の自由度が低いため、両者を併用しても時空間全体について粒度の小さなスケーラビリティを実現することはできない。小文では不均一フレーム間隔によって時間スケーラビリティを実現する手法を提案し時間方向の粒度の精細化を図ると共に、FGS と併用して主観品質を最大化するスケーラブル符号化を実現するための基礎検討を行なう。

## Scalable Video Coding based on Subjective quality — An Approach of Nonuniform Frame Period —

Takamichi MIYATA, Takenobu USUI and Toshiyuki YOSHIDA<sup>†</sup>

Dept. Communications and Integrated systems

Graduate school of Science and Technology

Tokyo Institute of Technology

2-12-1 O-okayama, Meguro-ku, Tokyo 152-8552, Japan

†Dept. Information Science

Fukui University

3-9-1 Bunkyo, Fukui 910-8507, Japan

E-mail : miyata@ip.ss.titech.ac.jp

**Abstract** While the Fine Granular Scalability (FGS), which has been proposed as one of SNR scalability techniques, offers improved freedom for bit rate control, a combination of the FGS and a conventional temporal scalability cannot realize a scalable coding scheme with fine granularity, because the frame rate selection in the conventional temporal scalability technique is inevitably restricted. This paper tries to realize a fine-granular temporal scalability by introducing a nonuniform frame period, and proposes a fundamental technique for maximizing subjective quality of scalable-coded videos by making use of the temporal scalability and the FGS together.

# 1 はじめに

近年、高速・大容量ネットワークの整備に伴い、動画像伝送を利用した様々なアプリケーションが急速に普及している。しかし、インターネットやモバイル端末などの帯域保証の無いに環境においては、ユーザの利用可能な帯域は時間的に変動するため、単一のビットレートで圧縮された符号化動画像のみを用いて動画像伝送を行う場合は、帯域の低下に伴ってサービス品質 (QoS: Quality of Service) の低下が起こる。そのため、ビットストリームを低帯域用の基本レイヤと画質向上用の拡張レイヤに分けるスケーラブル符号化が提案されている。

一方、著者らは、動画像の主観品質が単一フレームの品質 (空間方向の品質) と動きの滑らかさ (時間方向の品質) の両者から決定されることに着目し、与えられたビットレートにおいて主観評価値 (MOS: Mean Opinion Score) を最大化する最適フレームレート (Frame Per Second (FPS)) の推定手法を提案している [1]。さらに、これをスケーラブル符号化に応用した MOS スケーラブル符号化法の検討も行っている [2]。しかし、この手法では、MPEG-2 の SNR-時間ハイブリッドスケーラブル符号化を用いているため、空間方向 (SNR) のスケーラビリティの粒度が粗い問題をはじめとして、いくつかの未解決の問題が残されていた。

これに対し、近年、MPEG-4 FGS (Fine Granular Scalability) と呼ばれるスケーラブル符号化法が提案され [3]、MPEG-4 Streaming Video Profile として規格化されている。この手法では、拡張レイヤのビットストリームを任意の位置で切り出しても SNR の向上が可能であり、これによってユーザの利用可能な帯域に応じた QoS を提供することができる。さらに、FGST と呼ばれる SNR-時間ハイブリッドスケーラビリティ符号化も提案されており、時間方向と空間方向のレートの分配についても検討が行われている [4]。このビット量配分手法は PSNR に基づいているが、このような制御は本来、人間が知覚する主観品質に基づいて行われるべきである。

さらに、文献 [2]、[4] の手法に共通する問題点として、時間スケーラビリティにおける粒度の粗さが挙げられる。これらの手法では、図 1 に示すように、基本レイヤのフレーム間に、拡張レイヤのフレームを一樣な間隔で挿入しているため、基本レイヤ、拡張レイヤでの FPS をそれぞれ  $t_b$ 、 $t_e$  [frame/sec] とすると、例えば  $t_b = 5$  に対しては  $t_e = 10, 15, 30$  の 3 種類に限定され、また  $t_b = 10$  に対しては  $t_e = 30$  のみに限定される。

以上の背景の下、本稿では、前述の最適フレームレート推定法と、不均一なフレーム間隔を FGS 符号化に応用することにより、時空間の両方向について、粒度の細かい主観品質主導型のスケーラブル符号化を実現するためのアプローチについて述べる。

以下、まず 2. では以前に提案した最適フレーム

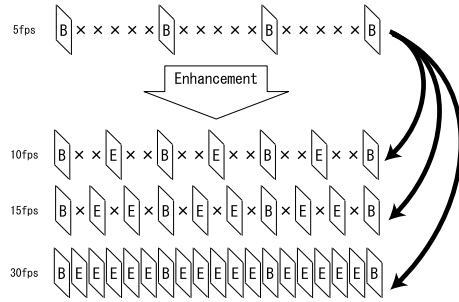


図 1:  $t_b = 5$  に対して取り得る  $t_e$

レート推定と、これに基づくスケーラブル符号化法について述べ、この手法の問題点について述べる。次に、3. では、空間方向の主観品質の予測に対して PSNR を用いるアプローチを提案し、4. では時間方向の粒度の細かいスケーラビリティを実現するための、不均一フレーム間隔によるアプローチについて述べる。5. では提案する主観品質主導型の FGS 符号化のレート制御アルゴリズムの概要を示す。最後に 6. でまとめを行う。

## 2 最適 FPS の推定とそれに基づく MOS スケーラブル符号化法

動画像を符号化する際のビットレートを  $c$  [kbit/sec] とすると、 $c$  は、

- 1 秒あたりのフレーム数 (Frame Per Second (FPS))  $t$  [frame/sec]
- 1 フレームあたりの平均ビット量 (Bit Per Frame (BPF))  $s$  [kbit/frame]

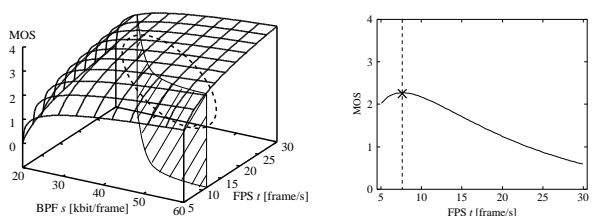
を用いて  $c = st$  と表すことができる。いま、空間解像度が一定であるとすれば、符号化動画像の品質は、この FPS と BPF によって決まる。前者は動きの滑らかさに相当する「時間方向の品質」を、後者は単一フレームの画質である「空間方向の品質」をそれぞれ決定するため、動画像の符号化においては、この両者をシーンに応じてバランス良く制御する必要がある。

### 2.1 文献 [1] における最適 FPS 推定法

文献 [1] の手法は、BPF  $s$  [kbit/frame] と FPS  $t$  [frame/sec] に対する符号化対象動画像の MOS の分布を表す関数  $MOS(s, t)$  を予測することに基づいている。

文献 [1] では、最適フレームレートを以下の手順で求めている。

1. 符号化対象の動画像から抽出した特徴量を用いて、 $MOS(s, t)$  を予測する (図 2(a))
2.  $c = st$  を満たす双曲線と、 $MOS(s, t)$  との交線を BPF 軸方向に投影し (図 2(b)),  $MOS(c/t, t)$  を最大にする  $t$  (図中の×) を最適フレームレートとする

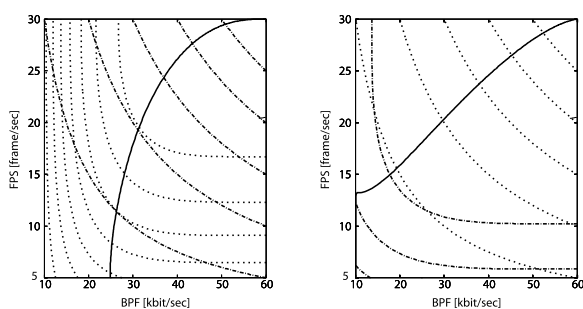


(a) 品質の時空間特性 (b) FPS と MOS の関係

図 2: 最適フレームレートの推定

ここで、符号化動画像の品質は、符号化器およびそのパラメータによって決まるため、文献 [1] では、 $N = 12$  および  $M = 3$  の固定 GOP 構造をもつ MPEG 符号化器を対象として最適 FPS 推定法を与えている。

このようにして求めた 2 種類の動画像の最適 FPS の軌跡の例を図 3(a), (b) に示す。図中の破線は  $MOS = \text{一定}$  の等高線、一点鎖線は  $c = \text{一定}$  の等高線、実線は最適 FPS の軌跡を表している。すなわち、与えられたビットレートを表す双曲線と、当該曲線の交点が最適フレームレートを表している。この図から、 $MOS$  が最大となる FPS と BPF の組み合わせが、動画像の内容に強く依存していることが分かる。



(a) 画像 a (b) 画像 b

図 3: 最適 FPS の軌跡

## 2.2 最適 FPS 推定法を用いた MOS スケーラブル符号化

SNR-時間スケーラブル符号化では、ビットレートを基本レイヤと、基本レイヤの SNR および/もしくは FPS を向上させるための一つないしは複数の拡張レイヤに分割し、ユーザの利用可能な帯域

に応じて異なる QoS を提供する方法である。この手法では符号化時に、拡張レイヤのビットレートを時間方向 (FPS) と空間方向 (BPF) にどのように分配するかを決定する必要がある。ここでは、文献 [1] の最適 FPS 推定法を用いて MOS を最大化する MOS スケーラブル符号化について考える。

基本レイヤと拡張レイヤの符号化特性が等しいと仮定すれば、文献 [1] の最適 FPS 推定を用いて求められた最適 FPS の軌跡 (図 3) を用いることで、基本レイヤと拡張レイヤにおける最適フレームレートが容易に得られる。しかし、一般に SNR スケーラビリティでは基本レイヤに比べてオーバーヘッドが発生し、時間スケーラビリティでは GOP 構造 (B ピクチャの枚数) が変化するため、符号化効率が変わる。文献 [1] の手法では、空間方向の主観評価値の予測に BPF と MOS との関係を用いているため、拡張レイヤにおける符号化効率の変化に対して何らかの対策が必要となる。

文献 [2] においては、近似式を用いて符号化効率の変化を推定していたが、この式の妥当性については十分に検証されていない。また、前述のように取り得る  $t_e$  が  $t_b$  に強く依存するため、必ずしも求められた  $t_e$  に従って符号化が行えるとは限らない。いま、 $t_b = 5$  であるとする、図 4 から、 $t_e = 7.5$  を選ぶことはできず、実際に取り得る  $t_e$  は 10, 15, 30 に限定される。そのため、時間方向のスケーラビリティの粒度は粗いものとなる。

また、文献 [2] が仮定しているスケーラブル符号化法では、空間方向 (SNR) におけるスケーラビリティの粒度も粗いものとなる。

文献 [2] の MOS スケーラブル符号化に関する課題を整理すると、以下ようになる。

- 符号化効率の変化に柔軟に対応できない
- 時間方向のスケーラビリティの粒度が粗い
- 空間方向のスケーラビリティの粒度が粗い

以下では、これらの課題についてのアプローチについて述べる。

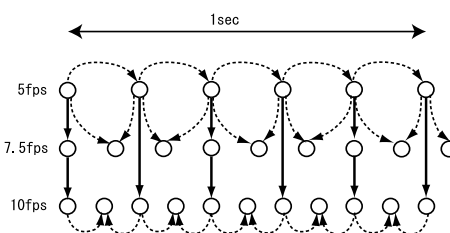


図 4: FPS の拡張時の問題点

### 3 PSNRに基づく主観品質の推定

著者らはこれまで、空間方向の主観評価値を予測するために、BPFとMOSとの関係を主観評価実験の結果に基づいて画像の特徴量から推定する方法を提案してきた[1], [2]。しかし、この方法では、GOP構造や動き予測の探索範囲などの符号化器のパラメータの変化などによって符号化特性が変わると、そのたびに主観評価実験を行う必要が生じる。特にスケーラブル符号化においては、拡張レイヤでの符号化特性が基本レイヤのビットレートに応じて変化してしまうため、実用的とはいえない。

一方、客観評価値であるPSNRとMOSの間には、単一の画像および劣化要因に対しては高い相関があることが知られている（ただし、一定のMOSを与えるPSNRの値が画像によって大きく変化することも事実である）。

そこで、BPFとMOSの関係を求める際にPSNRを経由し、BPFとMOSの関係を、BPFとPSNR及び、PSNRとMOSの2段階に分けて考える。BPFとMOSの関係は符号化器パラメータによって変化するが、PSNRとMOSの関係は符号化器パラメータにはほとんど依存せず、画像の性質にのみ依存すると考えられる。したがって、主観評価実験は、一つのエンコーダに対して一度行えば良く、スケーラブル符号化等への応用が容易となる。

以下では、対象動画画像から抽出した特徴量を用いて、PSNR-MOS特性を推定する手法について述べる。

#### 3.1 主観評価実験の方法

PSNR-MOS特性を推定する手法を構築するうえで、教師データとなるMOSを得るための主観評価実験を行った。実験は、6人の被験者に2回ずつのべ12人で行なった。被験者にはITUの勧告に基づくトレーニングを行った後、異なる平均PSNRを有するサンプル動画画像を提示し、0~5の6段階で評価させた。実験には、8種類の動画画像（すべてFPS=30）をMPEG-2によって圧縮したものを使用した。

#### 3.2 主観評価実験の結果

主観評価実験を行なった結果を図5に示す。PSNRとMOSとの関係には、次のような特徴が見られる。

- 各動画画像における傾きはほぼ等しい
- 一定のMOSを与えるPSNRの値が高い動画画像ほど、直線に近くなる傾向がある

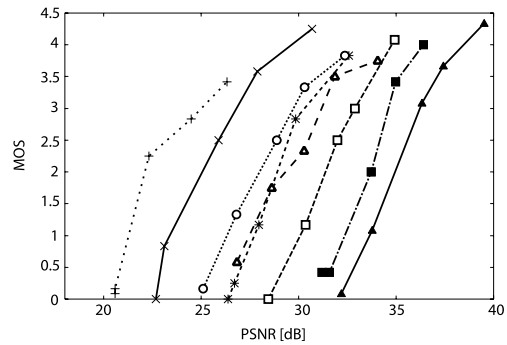


図5: 主観評価実験の結果 (PSNR-MOS の関係)

#### 3.3 PSNR-MOS 曲線の推定

PSNR-MOS 曲線を近似する関数として、

$$MOS = \alpha\tau(x - \tau)^3 + \beta\tau(x - \tau) + \gamma \quad (1)$$

を用いる。ここで、 $x$ はPSNRの値で、 $\alpha$ 、 $\beta$ 、 $\gamma$ は定数、 $\tau$ は動画画像に依存する変数である。また、 $\gamma$ はPSNRが $x = \tau$ におけるMOSの値に対応し、ここでは $\gamma = 4$ とにおいて、 $MOS = 4$ を与える $\tau$ の値を各画像に依存する変数とする。

最小自乗法を用いて各動画画像に対する $\alpha$ 、 $\beta$ 、 $\tau$ の値を求めた後、 $\alpha$ 、 $\beta$ の値を全ての画像で平均して定数 $\alpha$ 、 $\beta$ の値を決定する。この $\alpha$ 、 $\beta$ を用いて、再度最小自乗法を行い、各動画画像ごとに $\tau$ の値を求めた。

#### 3.4 回帰分析

次に動画画像の特徴量を用いて $\tau$ の値を推定し、対象動画画像固有の推定曲線を求める。一般に、エッジが多い画像ほど符号化に伴う劣化が目立ち難く、低いPSNRで高いMOS値を示す傾向があるため、エッジの統計量から $\tau$ の値を推定する。具体的にはSobelフィルタの出力を全画素で平均したものを特徴量として用いた。

計測した8種類の画像のうち、2種類をopen data、残りの6種類をclosed dataとして、open dataに対するPSNR-MOS特性を推定した結果を図6に示す。MOSが2以上の領域では誤差が0.6程度と良好に推定できているが、MOS=1以下では、1以上の誤差が生じている。しかし、実際に最適フレームレート推定に応用する場合には、MOSが2付近より上で正確に推定可能であれば良いため、MOS=1以下における誤差は実用上ほとんど問題にならないと考えられる。

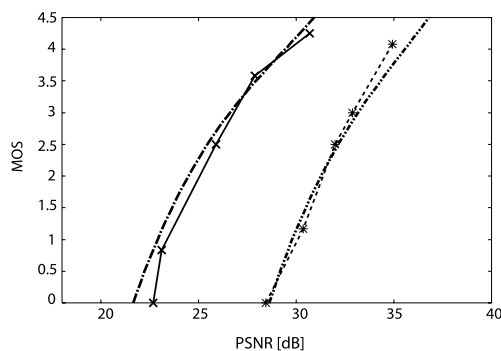


図 6: 計測結果及び推定した MOS-PSNR 曲線 (open data)

## 4 不均一フレーム間隔の導入

文献[1]の最適FPS推定法では、時間方向の品質  $MOS_t$  を、FPS  $t$  の関数  $MOS_t(t)$  として画像の特徴量から推定していた。これはすなわち、一定の時間 (GOP や単一シーンなど) 内において、フレーム間隔が均一であるという条件に従っていたことになる。そのため、MOS スケラブル符号化を行う際には、時間方向の粒度が粗くなり、最適 FPS の軌跡に沿った符号化が困難となっていた。一方、動画の時間方向の品質劣化の種類として、動画のフレーム間に発生する動きの不自然さ (ジャークネス等) が挙げられる。この動きの不自然さと、動画のフレーム間での動きとは強い関係があるため、図7のように、フレーム間での動きに応じてフレームを挿入することで、効率的に、かつ細かく時間解像度を変化させることができる。

### 4.1 時間方向の品質の推定式の変更

動画の時間方向の品質は、動きベクトルの統計量  $m$  の最大値  $m_{max}$  と強い関係がある。ここで、その統計量として、各マクロブロックの動きベクトルに対して信頼度 [5] の低いものを除いた上で、全マクロブロックでの最大値を取った。これらの結果から、 $m_{max}$  を様々に変化させた動画について主観評価実験を行い、推定された時間方向の品質  $MOS_t$  を新たに  $m_{max}$  の関数  $MOS_t(m_{max})$  として表す。この  $m_{max}$  に対する曲線  $MOS_t(m_{max})$  を画像の特徴量から推定し、FGST のレート制御に用いることで、時間・空間方向の両方について粒度の細かいスケラブル符号化が可能となる。実際には、図8に示すようにフレームを追加することで  $m_{max}$  が小さくなり、MOS を向上させることができる。

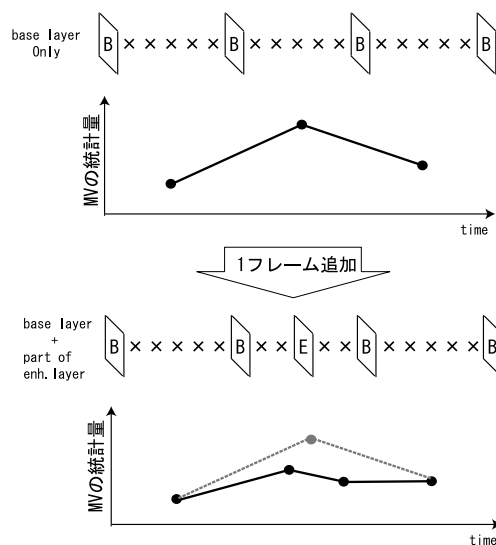


図 7: 不均一フレームレートにおけるフレーム追加の様子

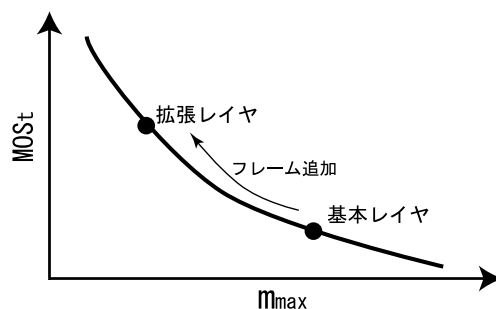


図 8: 不均一フレーム間隔による時間スケラビリティ

## 5 主観品質主導型のスケラブル符号化法

本節では、これまでに述べてきた最適 FPS 推定、PSNR に基づく主観品質の推定、不均一フレーム間隔による時間解像度の制御法を FGST に応用し、主観品質主導型のスケラブル符号化法の概要を提案する。この手法を用いることで、ユーザの利用可能な帯域に応じ、時間方向と空間方向の双方において細かい QoS 制御が可能となる。ここで、拡張レイヤにおいては、符号化方式の違いから符号化効率が変わるため、空間方向の主観品質の推定には前述の PSNR に基づく手法を用いる。

いま、基本レイヤ (以下、 $b$ )、拡張レイヤ (以下、 $e$ ) のビットレートをそれぞれ  $c_b, c_e$  [kbit/sec] とし、 $c = c_b + c_e$  とする。 $b, e$  の BPF をそれぞれ、 $s_b,$

$s_e$  [kbit/frame] とし,  $b$  の FPS を  $t_b$  とする. 文献 [1] の手法によって,  $s_b$ ,  $t_b$  を決定できる.

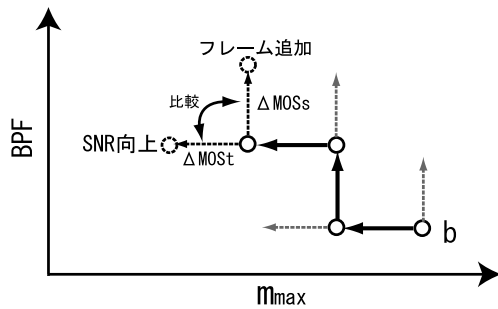


図 9: MOS を最大化するフレーム追加数の推定法

まず, 基本レイヤでの  $m_{max}$  を求め,  $m_{max}$  となる  $m$  を持つフレーム間隔に, 拡張レイヤのフレームを 1 フレーム挿入することを考える. 挿入したフレームの PSNR は, 既存のフレームと同等である必要があるため, 挿入に必要なビット量  $c_d$  は BPF-PSNR 特性の予測曲線を用いるか, 実測を行えば一意に決まることになる. フレームの挿入により  $m_{max}$  が変化するため, MOS が向上する. このときの MOS の変化量の推定値を  $\Delta MOS_i$  とする. また,  $c_d$  をフレームの挿入ではなく, 既存のフレームの SNR の向上に使用したときの MOS の増分  $\Delta MOS_s$  を, BPF-PSNR-MOS の関係を用いて計算する.  $\Delta MOS_i$  と  $\Delta MOS_s$  を比較し,  $\Delta MOS_i$  が大きければフレームの追加を行い,  $\Delta MOS_s$  が大きければ, このビット量を SNR 向上に用いる. この手順を  $c_d$  の積算が  $c_e$  に到達するまで繰り返す. 図 9 にこの様子を示す.

これにより, 与えられたビットレート  $c_e$  において MOS を最大にするフレーム追加数が分かり, 最適な分配が決定する. また, この過程において, ユーザの使用できるビットレートが  $c$  よりも小さい場合においても, その軌跡から MOS を最大化する最適フレーム数が求められる.

## 6 まとめ

時空間の両方向について粒度が細かく, かつ主観品質によって主導されるスケーラブル符号化法の概要について述べた. これまでに提案してきた最適フレームレート推定に加え, PSNR を利用した空間方向の主観評価値の予測や, 不均一なフレーム間隔を FGST 符号化法に应用することで, 目的とする符号化法の実現可能性が示された. 今後の予定としては, 主観評価実験の実施と, それに基づくアルゴリズムの詳細の決定, さらに提案方式の実装が挙げられる.

## 参考文献

- [1] 稲積, 吉田, 酒井, 堀田: “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の推定”, 信学論, Vol.J85-B No.7 pp.1130-1142, (Jul. 2002).
- [2] 稲積, 吉田: “動画像伝送における MOS スケーラブル符号化の一検討”, SD-3-13, 電子情報通信学会総合大会, pp.397-398, (Mar. 2002).
- [3] W. Li, “Overview of fine granularity scalability in MPEG-4 video standard”, IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 3, pp. 301- 317 (Mar. 2001).
- [4] R. Kumar, M. van der Schaar, S.F. Chang: “FGS+: Optimizing The Joint SNR-Temporal Video Quality in MPEG-4 Fine Grained Scalable Coding,” Proc. IEEE ISCAS 2002, (Apr. 2002).
- [5] 吉田, 宮本, 酒井: “動画像の動きベクトルに対する信頼度関数とその応用”, 信学論 D-II, J80D-II, 5, pp.1192-1201, (May 1997).