

スケーラブル映像符号化および配信システムに関する検討

嵯峨田 淳 八島 由幸 上倉 一人 小林 直樹

NTT サイバースペース研究所

〒 239-0847 横須賀市光の丘 1-1

TEL: 0468-59-4673, FAX: 0468-59-2829

E-mail: sagata@nttvdt.hil.ntt.co.jp

本稿ではインターネット映像配信のためのスケーラブル映像符号化方式および配信システムについて提案する。提案符号化方式はウェーブレット変換および動き補償を伴う時間方向ハール変換を用いて符号化を行い、空間解像度・時間解像度が組み合わされた形式で符号化ストリームに記録される。このため中継サーバにおけるレート変換処理は共通の入力ストリーム内のヘッダを探索し以降のデータを取り除く簡易な処理で済むため、中継サーバは複数のクライアントに対して同時にレート変換処理を行い、変換されたストリームをそれぞれ利用者に配信する事が可能となる。

Scalable Video Coding and its Delivery System over the Internet

Atsushi SAGATA Yoshiyuki YASHIMA Kazuto KAMIKURA Naoki KOBAYASHI

NTT Cyber Space Laboratories

1-1 Hikarino-oka, Yokosuka, Kanagawa 239-0847, JAPAN

TEL: +81-468-59-4673, FAX: +81-468-59-2829

E-mail: sagata@nttvdt.hil.ntt.co.jp

This paper describes scalable video coding strategy and its implementation in a live delivery system over the Internet. This coding strategy possesses scalability with regard to frame rate, resolution, picture quality, and bitrate. Using its data structure, a relay server can easily reduce the bitrate and re-generate a new bitstream with only a trivial loss in coding efficiency and frame rate by truncating the input bitstream transmitted from an encoder or a parent server. The calculation cost of truncating the bitstream is also so negligible as to allow for simultaneous re-generation of multiple bitstreams at any bandwidth. Consequently, end users can receive the maximum level of video quality according to the given bandwidth.

1 はじめに

インターネットを介しての映像配信 [1] が注目されている。利用者からインターネットまでの回線は一般的に低ビットレートであり、また利用者の接続環境によりそれぞれ伝送レートが異なる。さらにインターネットは QoS 非保証のネットワークであるため、輻輳などの原因により実際に利用可能な伝送帯域が実時間で変動する [2]。このため、符号化データは効率良く圧縮するのみならず、とりわけ多地点への映像配信を行う際には、共通の符号化データから、利用可能な伝送帯域に応じた量のデータを取り出し復号できる技術が必要であると考えられる。

インターネット上での映像配信サービスは、大きく分けてリアルタイム系と非リアルタイム系のサービスに分類される。非リアルタイム系のサービスであれば、前述の問題点の解決策として、前もってストリーム(符号化データ)をいくつかのビットレートで符号化しておき、利用者の利用可能な帯域幅あるいは帯域の変動に合わせて配信するストリームを切替える方法 [3] が考えられる。しかしこの方法には、配信映像の一元管理ができず、ビットレートの異なる複数のストリームを並列して符号化するため膨大な演算時間を要するという問題がある。さらに、この方法をリアルタイム系のサービスに適用した場合には、異なる帯域幅

を有する複数のユーザに対して映像ストリームを配信するとき、符号化器からインターネットの基幹回線までの間の回線と同じ映像内容でビットレートの異なるストリームが流れる事となるため、この回線がボトルネックとなり非効率である。したがって、リアルタイム系のサービスの場合は、ある共通の単一符号化ストリームから、利用可能な伝送帯域に応じた量のデータを取り出し、新しい符号化ストリームを生成する技術が今後不可欠になってくると予想される。

2 システム概要

2.1 映像配信システムの概要

1章に述べたように、リアルタイム系の映像配信サービスでは、利用者に配信する各ストリームは単一ストリームから生成されるのが望ましい。さらに、サーバからインターネットの基幹回線までの間の回線で同一映像内容のストリームが輻輳する問題を解決するためには、ネットワークに散在する中継サーバにおいて、入力ストリームを複数の異なるビットレートのストリームにレート変換し、ネットワーク下流に存在する中継サーバあるいは利用者にそれぞれ配信する方法が最適と言える。

スケーラブル映像配信システム の概念図を 図 1 に示す。中継サーバ RS_a は、ネットワークの上流に位置する符号化器より、 a kbps の符号化ストリーム A を受け取る。この符号化ストリームの基本レイヤを $a1$ kbps とし、拡張レイヤを $(a-a1)$ kbps とする。このときネットワーク内の中継サーバ RS_a では、符号化ストリーム A に対して以下のような処理が行われる。 RS_a から RS_b への伝送に関しては、伝送帯域が b kbps ($> a$ kbps) であるため、符号化ストリーム A が符号化ストリーム B としてそのまま転送される。一方、 RS_a から RS_c への伝送に関しては、伝送帯域が c kbps であり、符号化ストリーム A の全データは伝送できないが、基本レイヤのビットレートより帯域幅が広い。この場合 RS_a でレート変換が行われ、符号化ストリーム A の基本レイヤと拡張レイヤの一部

分が符号化ストリーム C として RS_c に伝送される。

2.2 符号化ストリームへの要求条件

多数の利用者が同時に存在しうるリアルタイム系のサービスでは、中継サーバでのレート変換処理は低計算量でなくてはならない。レート変換処理の計算量を少なくするためには、入力ストリームとして H.261 のような汎用的な符号化アルゴリズムを採用して、中継サーバでトランスコーディングのような計算量の大きい処理を行う方法は現実的とは言えない。この問題を解決するためには、符号化器から出力される共通の符号化ストリーム自体が基本レイヤ (base layer) と拡張レイヤ (enhancement layer) から構成され、拡張レイヤの中身がプログレッシブなデータ構造とされている必要がある。このようなスケラブルなデータ構造を有することにより、中継サーバは例えば単純に符号化ビット列を遮断するのみの低計算量のレート変換器として作動する事が可能となるため、結果として複数のクライアントに対して出力する符号化ストリームを同時に生成できるからである。

また、基本レイヤの符号化レート未満の符号化レートに変換する必要も考えられる。 RS_a から利用者 $U_{d,e}$ に伝送する場合には、伝送帯域が符号化ストリーム A の基本レイヤより狭い。このときレート変換後のストリームの品質を保ちたいのであれば、確かにトランスコーディングや再符号化のような処理が望ましい。しかし図 1 の例の場合、 RS_a には 2 つの中継サーバと 2 人の利用者が下流に直結されているため、これらに適切なデータ量の符号化ストリームを伝送するには同時に 4 並列のレート変換処理をする必要がある。このため、基本レイヤのデータに関しても、低計算量でレート変換できる方式が望ましい。

2.3 スケーラブルアルゴリズム

非スケーラブルなストリームを配信する従来のシステム [1] では、中継サーバにおいて最後の P ピクチャから順に落ちていくため、レート変換後のストリームのフレームレートが一定とならず、復号映像が動いては止まるなど、ぎくしゃくとした動きとなる問題が指摘されている。また、384~64 kbps 程度の伝送帯域幅を主なターゲットとするインターネット映像サービスの場合、ビットレートによって適切なフレームレートが異なる。このため MPEG2[4] で採用されている幾つかのスケーラビリティのように、レート変換処理によりフレームレートは変動せずに画品質のみが劣化する方式のみを採用した場合、伝送帯域幅に適した品質の映像を取り出すことは困難である。このため低ビットレートにレート変換処理を行った場合

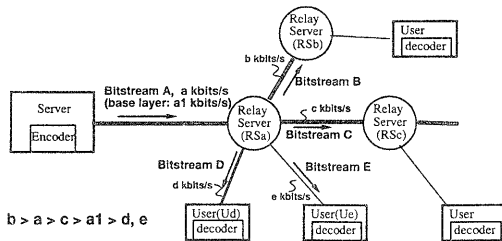


図 1: Video delivery system overview.

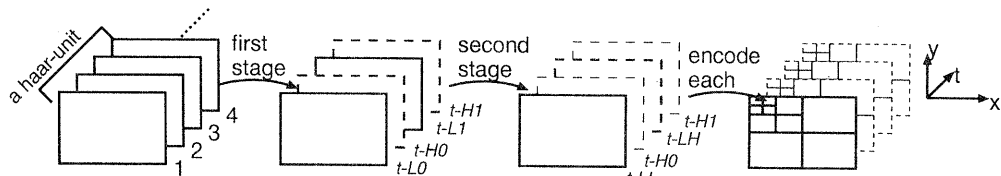


図 3: Generation of temporal subband.

は、元の共通ストリームと比較して画品質を劣化させるだけではなく、同時にフレームレートも均一に低下させることにより、見た目に自然な復号映像を生成する必要がある。

2.3.1 時間スケラビリティ

フレームレートを均一に変動させるためには、幾つかのフレームは予測符号化の際に参照されてはならない。本システムでは時間スケラビリティの実現のため、フレーム間予測符号化に動き補償を伴う時間ハール変換 (MC-TSB) [5] を採用した。図 2 に動き補償時間ハール変換の概要を示す。動き補償時間ハール変換では隣接する 2 枚の入力フレームにおいて、まず時間的に後の第 2 フレームに対してブロック分割を行い、各ブロックごとに第 1 フレームに対して動き予測を行う。このとき、対応する両ブロック内の各画素についてそれぞれハール変換を行う。ブロック分割を行った第 2 フレームが時間高域サブバンド ($b'=[b-a]/2$)、参照フレームである第 1 フレームが時間低域サブバンド ($a'=[a+b]/2$) となる。時間低域サブバンド内の参照されなかった画素領域はイントラ領域とし、第 1 フレーム内の対応する位置の画素に置換 ($a'=a$) される。時間ハール変換は動きベクトルに沿って行われるため、時間低域サブバンドは両入力フレームの動きを考慮した平均画像となる。以上の処理に加えて、第 1 フレーム内において二つ以上のブロックから参照される画素領域の処理 [6] を適切に行うことにより、動き補償時間ハール変換は一画素精度の完全再構成が保証される。

次に、時間サブバンドの生成方法の概要を図 3 に示す。第 1 ステップとして先に示した動き補償時間

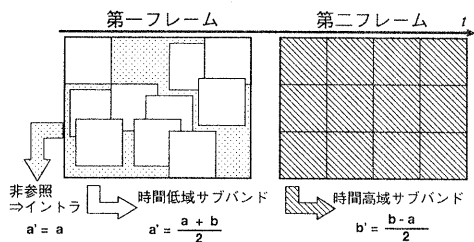


図 2: MC'ed temporal haar wavelet.

ハール変換を行うことにより、第 1, 2 フレームから $t(temporal)$ -L0 および t -H0 が、第 3, 4 フレームから t -L1 および t -H1 が生成される。次に第 2 ステップとして、同じ処理を t -L0 および t -L1 に対して行うことにより t -LL および t -LH が生成される。結果として得られる時間サブバンドは t -LL, t -LH, t -H0, t -H1 の 4 サブバンドとなる。以上の再帰処理は 3 回以上繰り返して行う事も可能であるが、動き補償の効率と符号化遅延を考慮して本稿での再帰処理を 2 回とし、入力 4 フレームを一符号化単位 (ハール符号化単位) として処理した。

2.3.2 解像度スケラビリティ

解像度スケラビリティ機能を実現するためには、各時間サブバンドとりわけ t -LL の符号化データの途中で打ち切られた場合でもそれなりの映像品質で復号可能でなければならない。このため、時間低域サブバンドおよび各時間高域サブバンドはウェーブレット変換 [7] を用い、変換係数を量子化後に低周波係数から高周波係数の順に符号化する。変換係数の符号化には既に提案済みである二次元可変長符号化 [8] を用いた。この方法では、子供係数がゼロツリールト (ZTR) である個数と符号化対象係数のレベルの絶対値の相関を用いて符号化を行う。この方法については 3 章に述べる。

2.3.3 画品質スケラビリティ

利用者のネットワークへの接続形態が多様な場合、あるいは利用者の利用可能な伝送帯域が実時間で大幅に変動する場合などには、時間・解像度スケラビリティを併用しても十分なレート可変機能を提供できない場合が考えられる。提案する映像配信システムではこれら二つのスケラビリティに加え、画品質スケラビリティを採用した。

本稿での画品質スケラビリティは符号化方式としては SNR スケラビリティ [4] に似た手法を採用している。SNR スケラビリティとは、基本レイヤで DCT 係数を量子化して符号化した結果として復号画像に内在する量子化誤差を、拡張レイヤで低減させる方式である。すなわち高 SNR 階層の量子化ステップ

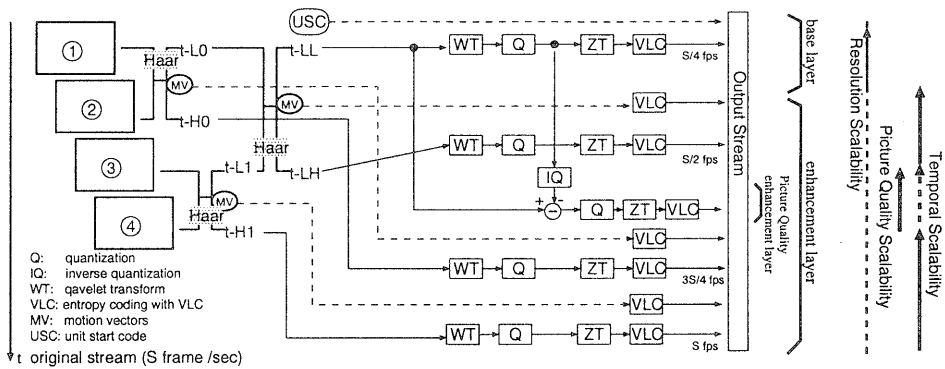


図 4: Scalable multiplexed syntax.

を低 SNR 階層より細かく符号化することで量子化誤差低減を実現する。画品質スケーラビリティでは、時間ハール変換により得られた $t-LL$ 時間サブバンドの量子化誤差を拡張レイヤで符号化する。すなわち基本レイヤあるいは直前の拡張レイヤの量子化ステップ値より細かい量子化ステップ値で、 $t-LL$ 時間サブバンドの量子化誤差を階層的に符号化する。

2.3.1項で述べたように、 $t-LL$ 時間サブバンドは動きを考慮した 4 フレームの平均画像である。したがって、 $t-LL$ 時間サブバンドの量子化誤差を符号化することにより、3つの時間高域サブバンドの量子化誤差を符号化する事なく、同一ハール符号化単位内の 4 フレームの画品質を改善する事ができる。

2.4 スケーラビリティ多重化

中継サーバにおいて簡易にビットレートを削減することを可能にする、符号化ストリームの構成のことを、ビットレートスケーラビリティと呼ぶ。ビットレートスケーラビリティは、2.3節に述べた3つのスケーラビリティを符号化ストリーム中で効率良く組み合わせる事により実現される。

提案方式における、 $t-\{H1, H0, LH, LL\}$ の各時間サブバンドと $t-LL$ 時間サブバンドの拡張レイヤを符号化する際の符号化シンタックスを図 4 示す。図中の出力ストリーム (Output Stream) への矢印は、各可変長符号化部から出力ストリームへの符号化データの出力順を表しており、各ハール符号化単位の先頭を表す符号 (Unit Start Code: 17 bits) に続いて、上から順番に符号化ストリームに出力される。

図右の上向きの矢印は、中継サーバにおいて符号化ストリームのある部分が遮断された際にどのスケーラビリティが利用されるかを表している。本シンタックスではまず時間スケーラビリティ機能が利用される。すなわち $t-H1$ および $t-H0$ 時間サブバンドの符号化データが削除されることにより、レート変換後の符

号化ストリームのフレームレートがレート変換前の半分 ($S/2$ kbps) にまで徐々に低下する。 $t-H0$ および $t-H1$ 時間サブバンドの符号化データの削除はフレームレートの低下と同時に、残りの第 1, 3 フレームの画品質も徐々に低下させる。この点で時間ハール変換によるスケーラビリティ機能は、フレームレートだけではなく、同時に出力画像の品質もスケーリングすることとなる。

中継サーバにおいて更にデータが削除された場合、画品質スケーラビリティ機能が利用される。画品質スケーラビリティは、拡張レイヤにおいて $t-LL$ 時間サブバンドの量子化誤差を符号化したものであるが、中継サーバにおいてこのデータを削除することにより、同じハール符号化単位内の 4 フレームの画品質が徐々に低下される。そして、さらに符号化データの削除を続けると $t-LH$ 時間サブバンドの符号化データが削除され、全ての高域時間サブバンドが削除された時点で最終的にフレームレートは元の $1/4$ に漸近する。以上が拡張レイヤを利用したレート変換である。

中継サーバから利用者までの伝送帯域が非常に狭く、以上のスケーラビリティの利用でも不十分な場合は、基本レイヤ内の符号である $t-LL$ 時間サブバンドの高周波成分を削除することによりさらなる符号量削減を行う。ただし、隣接する $t-LL$ 時間サブバンド間には予測符号化が適用されているため、以降の $t-LL$ 時間サブバンド間の復号時にドリフトに起因する歪みが生じることとなる。

3 エントロピー符号化

3.1 従来方法の問題点概要

ゼロツリー符号化などの多くのウェーブレットを用いた符号化方式では、対応する周波数帯域間で相似な場所の係数間にツリーを生成し、これを低周波係数から高周波係数にかけて順に符号化するため、解像度

スケーラビリティの機能を有する。さらに代表的な符号化方式である EZW[9] では、有意係数がビットプレーン展開され、MSB より順に符号化されるため、画品質スケーラビリティ機能も利用することが可能となる。静止画像符号化では、解像度・画品質スケーラビリティを組み合わせた符号化データを生成する事により、復号器では受信済みの符号化データを用いて漸次復号し、時間の経過とともに段階的に映像を出力できる。一方、動画の符号化の場合は、変換係数をスカラ量子化したものを直接符号化 [10] する方式も多く提案されている。しかし、ゼロツリー符号化には以下の3つの問題点があると考えられる。

(a) 符号化体系の冗長性

(1) ある変換係数 C_0 のレベルがゼロであり (2) 子孫係数 $D(C_0)$ の中に一つ以上有意係数が存在するとき、 C_0 はIZ(孤立ゼロ) 符号として符号化される。このとき図5に示すように C_0 の直接の4つの子供係数を $O(C_0)$ と定義した場合、 $c \in O(C_0)$ が全て ZTR であることはあり得えないが、実際には $c \in O(C_0)$ が全て ZTR である符号化データを、復号器は復号可能である。これは符号体系の冗長性を示す。

(b) 絶対値と子供係数の相関を未利用

C_0 のレベル ($Q[C_0]$) がゼロである場合、子孫係数のレベル ($Q[c], c \in D(C_0)$) が全てゼロになる確率が高くなる特性を利用して、これを一つの ZTR 符号に置換して圧縮を行う。しかし実際には、レベル $Q[C_0]$ がゼロの場合だけではなく、レベルの絶対値 $|Q[C_0]|$ が小さい程、4つの子孫係数 $c \in O(C_0)$ を幹とするサブツリーがそれぞれ ZTR となる確率が高くなり、逆に $|Q[C_0]|$ が大きい程、 $c \in O(C_0)$ がそれぞれ ZTR となる確率が低くなる。この相関を利用していない。

(c) ZTR の出現確率が極めて高い

符号化ビットレートが低いとき、一般的に ZTR の出現確率が増加して 50% 以上となる。このため、可変長符号化の際の各シンボルの出現確率を $1/2^n$ で近

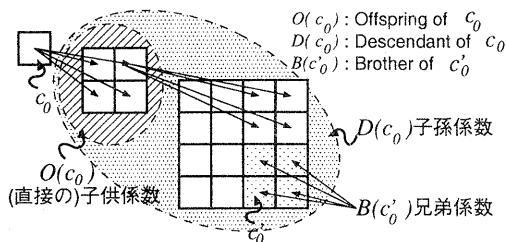


図 5: Definition of coefficient sets.

似することが困難となり、可変長符号の平均符号長が、そのエントロピーと比較して著しく増大する。

3.2 親子関係を用いた二次元可変長符号 [8]

解像度スケーラビリティを有しつつ、3.1節に述べた3つの問題を解決するため、ウェーブレット変換係数を低周波係数から高周波係数の順に符号化するにあたり、ある係数 C_0 を符号化する際に、子供係数 $O(C_0)$ を幹とするサブツリーが ZTR である個数 (ZC) と、 C_0 のレベル絶対値 $|Q[C_0]|$ からなる二次元可変長符号を用いてエントロピー符号化する。これにより、問題点 (b) に示した親係数のレベルと子供係数の ZTR の相関を効率的に利用することができる。

また、問題点 (a) を解決するため、 C_0 の符号化の際に C_0 の兄弟係数 $B(C_0)$ を一単位として走査し、連続して符号化データに出力する。 $B(C_0)$ 内の ZTR である係数は、親係数を符号化した際に出力済みの $B(C_0)$ の ZC と、兄弟係数内の ZTR のパターンを示す符号により、復号器側で一意に復号可能である。

4 計算機シミュレーション

4.1 シミュレーション条件

中継サーバに入力される共通のストリームを、標準画像 Foreman (QCIF, 30 fps, 300 フレーム) を用いて、ビットレート 256 kbps, フレームレート 7.5 fps にて生成した。フレーム間予測には動き補償を伴う時間方向ハル変換 [5] を2階層に施して4フレームを1ハル符号化単位として符号化し、隣接する $t-LL$ 間の動き予測にはオーバーラップ動き補償 [11] を用いた。各時間サブバンドに対して行うウェーブレット変換には5-3タップ整数フィルタ [7] を用いて周波数帯域分割を三回行い、変換係数にセンターデッドゾーン付きのスカラ量子化 [12] を行った後に、3.2節に示した二次元可変長符号を用いて符号化した。画品質スケーラビリティ機能を利用するため、各 $t-LL$ の拡張レイヤの量子化ステップをその基本レイヤでの半分の値に設定して符号化した。この画品質拡張レイヤの繰り返し回数は1回とし、基本レイヤの量子化ステップの半分の値で符号化した。

4.2 実験結果および効率検証

提案方式によるビットレートスケーラビリティの性能を確認するため、4.1節に示した条件に基づいて生成された共通のストリームを、中継サーバにおいて 240 ~ 10 kbps のレートのストリームに変換した。変換されたストリームの品質を図6に示す。図の横軸にレート変換後のビットレート (kbps), 左縦軸

にフレームレート (fps), 右縦軸に PSNR(dB) を示した。

256 kbps の入力ストリームのビットレートを削減することにより, まず始めに 2.4 節に示したように時間スケーラビリティが利用される。180 kbps のレートにレート変換するあたりでフレームレートが入力フレームレートの半分に漸近する。

次に 180 ~ 100 kbps のレートへの変換により, 画品質スケーラビリティが利用されてフレームレートの減衰が止まる。しかし, 図 6 の結果によるとこの区間のレートへの変換の際に, 画品質スケーラビリティが利用されているにも関わらず, 復号画像の PSNR がほとんど減少していない。この疑問は以下の理由によると考えられる。画品質拡張レイヤは平均画像である t -LL の品質を改善するための符号化データであり, 一方 PSNR は入力現画像と復号画像の相関値である。このとき t -L0, t -L1 とそれに対応する入力現画像 (第 1, 3 フレーム) は時間的に同じ位置ではあるが厳密には同じ画像ではないため, 画品質拡張レイヤの符号化データを削除しても PSNR の数値上はあまり変化が見られないと考えられる。事実, 180 kbps の復号画像は高周波成分が含まれた画像であり, 一方 100 kbps のストリームからはぼけた復号画像が再生されることから, 画品質拡張レイヤによる画品質スケーラビリティは正常に機能していると考えられる。

中継サーバにおいてさらにビットレートを削減すると, 40 kbps のレートに変換したあたりでフレームレートが $S/4$ に漸近し始める。すなわち, 図 4 に示した基本レイヤのビットレートが約 40 kbps であることを表す。これ以上のビットレート削減は, t -LL の高周波成分を削除することになる。隣接する t -LL 間には予測符号化が行われているため, ドリフトに起因する歪みが生じるものの, ウェーブレット変換およびオーバーラップ動き補償の特性によりブロック歪みなどの歪みは比較的知覚されにくい。

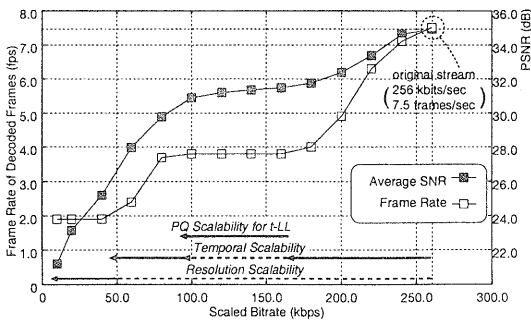


図 6: Simulation result of bitrate scalability.

5 まとめ

本稿では, 既に提案している二次元可変長符号化を用いて, スケーラブル映像配信のための符号化方式およびそのシステムを提案した。提案したスケーラブル符号化は, 時間・空間・解像度スケーラビリティを有し, これらを適切に組み合わせることにより, 極めて少ない演算量で中継サーバでのレート変換を実現した。現在実装を終えた映像配信システムでは, 中継サーバ間の実効帯域を実際には測定しておらず, 中継サーバでは手動で与えられたターゲットレートにレート変換した上で, 下流の複数のクライアントに映像を配信している。今後は中継サーバで実効帯域を実測しつつ配信する実験を続ける。

参考文献

- [1] 秦泉寺, 笠原: “インターネット用スケーラブル映像配信システム”, テレビジョン学会年次大会, 24-2, pp. 356-357, 1996.
- [2] 森西, 秦泉寺, 高田, 笠原: “インターネット上の実時間ストリーム配信に関する検討”, 信学ソ大, B-7-38, p116, 1997.
- [3] <http://www.real.com/>
- [4] ISO-IEC/JTC1/SC29/WG11: “Information technology - generic coding of moving picture and associated audio,” 13818-2, 1994.
- [5] J.-R. Ohm, “Temporal domain sub-band video coding with motion compensation,” Proc. International Conference on Acoustics, Speech and Signal Processing, Vol. 3, pp.229-232, Apr. 1992
- [6] S.-J. Choi and J. W. Woods, “Motion compensated 3-D subband coding of video,” IEEE Trans. Image Processing, Vol. 8, No. 2, pp. 155-167, Feb. 1999.
- [7] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” Communications on Pure and Applied Mathematics, vol.41, pp. 909-996, 1988
- [8] 嵯峨田, 八島, 小林: “ウェーブレット変換係数の二次元可変長符号化に関する一検討”, 情処学会研究報告, 2000-AVM-28, pp. 25-30, Mar. 2000
- [9] J. M. Shapiro, “Embedded image coding using zerotrees of wavelet coefficients,” IEEE Trans. Signal Processing, vol.41, no.12, pp. 3445-3462, Dec. 1993
- [10] S. A. Martucci et al. “A zerotree wavelet video coder,” IEEE Trans. Circuits and Systems for Video Technology, vol.7, no.1, pp.109-118, Feb. 1997
- [11] H. Watanabe and S. Singhal, “Windowed motion compensation,” Proc. SPIE conf. Visual Communications and Image Processing '91, Vol. 1605, pp.582-589, Nov. 1991
- [12] ITU-T Rec. H.263, “Video coding for low bitrate communication”, Dec. 1995.