

低レート向け動画像符号化のための 予測誤差解像度動的選択方式

伊藤 隆, 中川 章, 数井 君彦, 森松 映史, 松田 喜一
株式会社 富士通研究所

概要

PHS のデータ通信サービスや、既存のアナログ電話回線用の高速モデムを用いれば、数十 kb/s のデジタルデータ通信を行うことは容易であり、これを利用して動画像通信を行うこともできる。ところで、ITU-T で標準化された動画像符号化方式 H.263 は、多様な画像を少ない情報量に圧縮できる優れた方式である。しかし、動きが大きな部分では多くの情報が発生し、伝送できる画面の数が少なくなったり、個々の画面の品質が低下したりする。これら再生画質低下要因を取り除き、より高い品質の画像を低い伝送速度で提供するために、予測誤差信号の解像度を動的に切り替える方式を考案した。

Prediction Error Dynamic Resolution Scheme for Low Bitrate Video Coding

Takashi Itoh, Akira Nakagawa, Kimihiko Kazui, Eishi Morimatsu, and Kiichi Matsuda
FUJITSU LABORATORIES LTD.

ABSTRACT

High speed digital data transmission services have become available through PHS and 56 kb/s modems. Such services have ability to transmit motion pictures with certain coding algorithms. H.263 has been standardized by ITU-T, and is very powerful video coding algorithm for a wide range of input sequences. However, in highly active scenes, large amounts of bits are generated to encode each picture. This results in the frame skipping and the degradation of decoded picture quality. To keep these degradations as low as possible, we propose a mechanism of dynamic resolution switching of the prediction error. In the scheme, a reduced resolution coding is carried out when the input scene is highly active.

1. はじめに

PHS で提供されているデータ通信、あるいはアナログ電話網用の高速モデムにより、数十 kb/s のデジタルデータを送受することが容易になってきた。この通信メディアを利用するサービスの1つとして、動画像通信が挙げられる。しかし、動画像は 100 Mb/s を超える多量の情報を持つため、高性能な動画像符号化手法の導入が必須となる。

ところで、ITU-T で国際標準化された H.263 は、最新の研究成果を取り入れた優れた動画像符号化技術である。多様な画像を効率的に符号化することができる。しかし大きな動きのあるシーンでは、平均値を超える情報量の発生は避けられず、一定伝送速度の下では、単位時間に

伝送できる画面の数が減少して動きがギクシャクしたものになったり、あるいは各画面の品質が下がったりといった、サービス品質の低下が生じる。これは、H.263 だけでなく、それより前に標準化された H.261、あるいは ISO/IEC で標準化されている MPEG-1, -2, -4 といった方式が、フレーム間動き補償予測を採用していることによる。動きが大きいシーンでは、前提となっている画面間の相関の高さが満たされないため、符号化効率の点では望ましくないフレーム内予測が多く発生するからである。

この状況を改善するために、予測誤差解像度動的選択方式を考案した。これは、動画像中の動きの大きさを判定し、「大きい」と判断したときに予測誤差の解像度を水平垂直それぞれ

「MV」に対して擬似ベクトル値「pseudo-MV」を、

```
if MV == 0 then
  pseudo-MV = 0
else
```

```
  pseudo-MV = sign (MV) * (|MV| + 0.5) / 2
```

により定義し、この pseudo-MV に H.263 の動きベクトル用符号を割り当てる。

2.3. 縮小フィルタ

動き補償の後、 16×16 画素の予測誤差ブロックを 8×8 画素に縮小する。このときのフィルタ演算を簡易化するために、処理をブロックに閉じたものとした。具体的には図 3 に示すように、隣接する 4 画素の平均値を求めている。

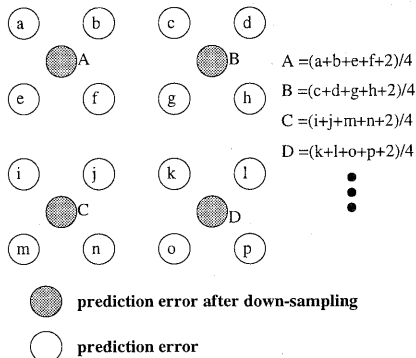


図3 縮小フィルタ

2.4. 拡大フィルタ

IDCT 演算の後、 8×8 画素の予測誤差ブロックを拡大して 16×16 画素の予測誤差ブロックを求める。各ブロック独立に処理できるようにするため、演算はブロックに閉じたものとなっている。図 4 に、4 つの画素 A, B, C, D から a, b, c, d を求める手法を示す。距離を反映して加重平均するものである。

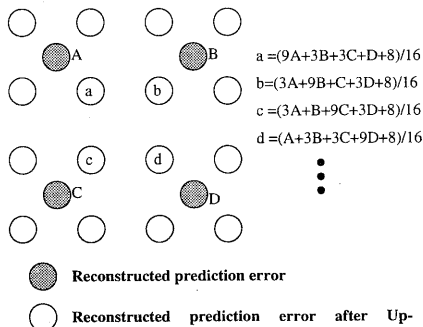


図4 拡大フィルタ

3. 切り替えアルゴリズム

3.1. 解像度決定

予測誤差の解像度は、フレームごとの「アクティビティ」に従って決める。すなわち、

■ 通常解像度から縮小解像度への切り替え

```
if ( $\overline{QP}_{i-1} \times B_{i-1} > TH1$ )
```

Switch to **reduced-resolution**;

■ 縮小解像度から通常解像度への切り替え

```
if ( $\overline{QP}_{i-1} \times B_{i-1} < TH2$ )
```

Switch to **normal resolution**;

ここで、 \overline{QP}_{i-1} は前の画面で用いられた量子化パラメータ QP の平均値、 B_{i-1} は前の画面の発生情報量を表す。

また閾値 TH1 は、通常解像度で再生画質の点から使用限界となる一番粗い量子化パラメータ「QP1」とフレームレート「FR1」とから、

$TH1 = QP1 \times (\text{設定ビットレート} / FR1)$

により計算する。

同様に TH2 は、縮小解像度で再生画質の点から使用限界となる一番良い量子化パラメータ「QP2」とフレームレート「FR2」とから、

$TH2 = QP2 \times (\text{設定ビットレート} / FR2)$

で求める。後に結果を示すシミュレーションでは、QP1=16, FR1=6, QP2=7, FR2=8 とした。

3.2. 縮小から通常への切り替え時の係数制限

縮小解像度に切り替わった後は、動いている部分の細かい絵柄は再生されない。したがって、通常解像度に戻るときは詳細情報を送らなければならないが、多量のデータが発生する。結果として、何枚もの画面がコマ落としされることになる。これを避けるために、縮小から通常に切り替わった直後の数画面の符号化では、DCT 係数に制限を加える。具体的には、切り替え後最初の通常解像度画面では、周波数の低い側の 4×4 個の係数、次の画面では 5×5 個、3 番目は 6×6 個、4 番目は 7×7 個を符号化対象とし、これらよりも高い周波数成分は切り捨てる。

4. シミュレーション

4.1. 条件

■ レート制御

低遅延の条件下で考案方式の有効性を評価するために、TMN8^[2]で規定されているレート制御方式を用いた。各画面の発生情報量が設定値になるような制御である。ただし、1 つの画面から大量の情報が発生した場合には、バッファ占有量を適度な値に保つた

め、コマ落としが起きることもある。

本シミュレーションでは、フレームレートを10とした。

■ 評価画像およびビットレート

標準化活動で用いられている "Mother & Daughter" や "Silent Voice" の他に、独自に作成した2種類の画像 "Bowing" および "Pamphlet" を評価画像として用いた。これらはいずれも、動きの大きな前景と、静止でかつ細かな絵柄がある背景を含んでいる。各画像の特徴と符号化ビットレートは、以下のようになっている。

- Bowing (48 kb/s)
人物が登場し、お辞儀をして退場する (図5参照)。
- Pamphlet (48 kb/s)
人物が登場し、パンフレットを取るために一度画面から消え、再度現れる (図6参照)。
- Mother & Daughter (48 kb/s)
母親と子供の上半身像。
- Silent Voice (48 kb/s)
上半身像であって、手話をしている。

4.2. 結果

図7～図10に、各画像での画面あたりのビット数 (各図左側上段のグラフ)、選択された解像度 (各図左側下段のグラフ)、および画質 (各図右側のグラフ) を示す。

図からわかるとおり、H.263 (各図の細線) では、各画面の発生情報量を目標値に合わせて一定に制御することは実現できていない。シーンによっては、目標値の3倍を超える情報が発生しているところもあり、コマ落ちは避けられない。一方提案方式では、情報量はよく制御できており、フレームレートをほとんど一定に保つことができている。これは、伝送できる画面数が多いというだけでなく、遅延時間にばらつきが生じないために動きが滑らかであるという点でも効果がある。また解像度切り替えの状況を見ると、H.263方式では規定の情報量に抑えられない部分で縮小解像度 (resolution low) が選択されていることがわかる。

画質 (PSNR) の面では、動きが大きくて縮小解像度が選択されている部分でも、提案方式と H.263 との差はほとんどないことがわかる。これと動きの滑らかさを合わせると、提案方式の方が主観品質で H.263 に勝るといえる。

5. まとめ

少ない情報量で動画像を伝送する場合に品

質向上を達成できる方式の1つとして、予測誤差解像度動的選択方式を考案した。動きが大きな画面に対して、予測誤差情報の解像度を下げて符号化する点、解像度が高い方に切り替わった直後の符号化係数に制限を設けて急激な発生情報量変化を抑えるようにした点、が特徴である。シミュレーションにより、既存標準方式である H.263 と比べて、動きの滑らかさを保ち、かつ各画面の品質も同等以上であることから、主観品質を向上させられることを示した。

6. 参考文献

- [1] "Temporal scalability algorithm based on image content", MPEG95/377
- [2] "Video Codec Test Model, Near-Term, Version 8 (TMN8) Release 0", Q15a59, ITU-T, Q15/WP3/SG16

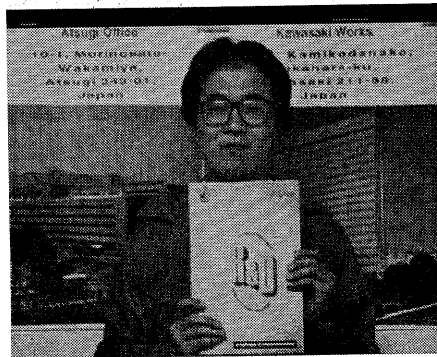
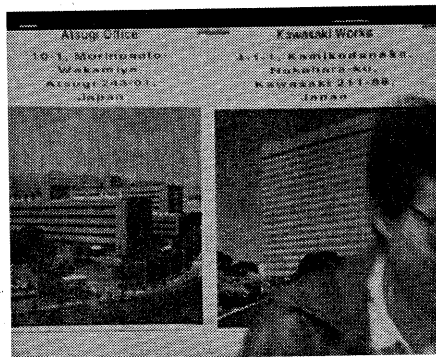
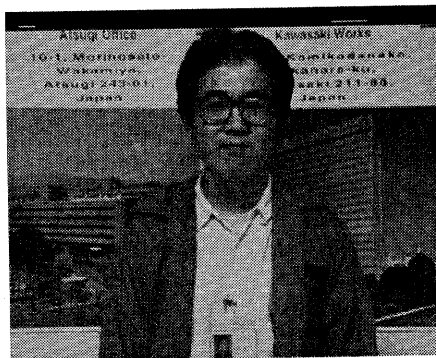
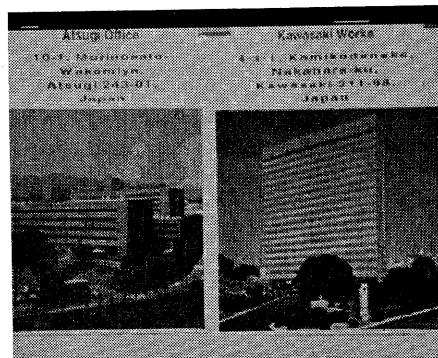
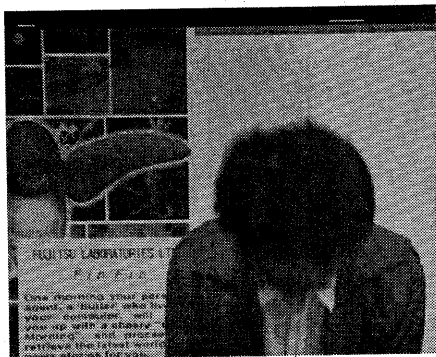
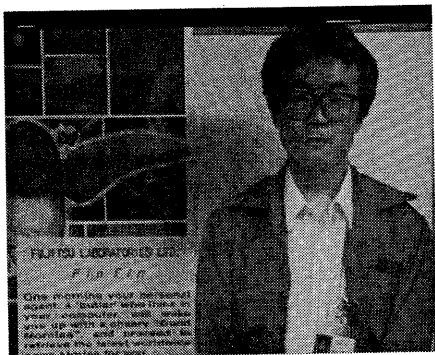
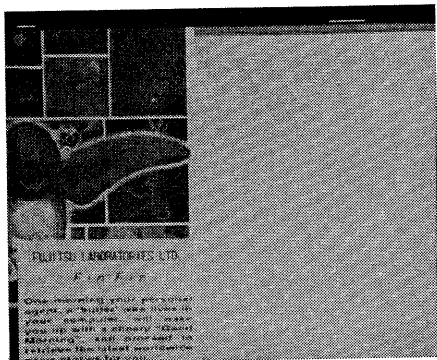


図5 画像"Bowling"の内容

図6 画像"Pamphlet"の内容

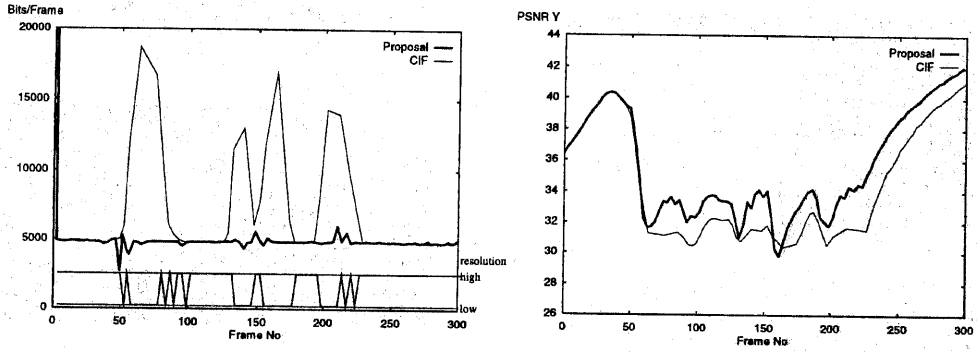


図7 "Bowing"の発生情報量, 解像度, 画質

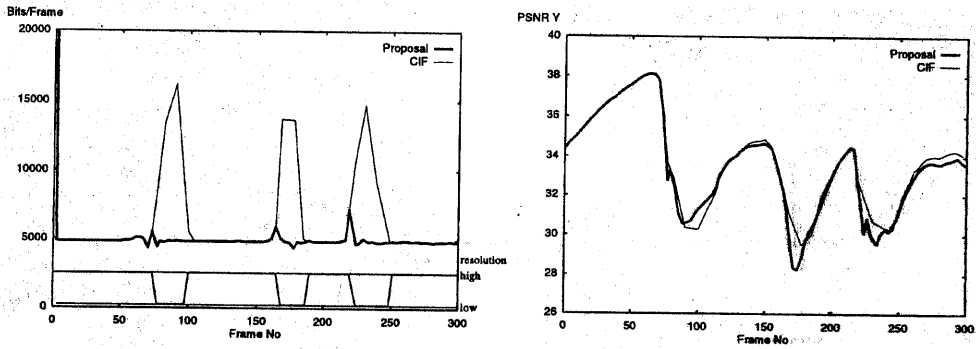


図8 "Pamphlet"の発生情報量, 解像度, 画質

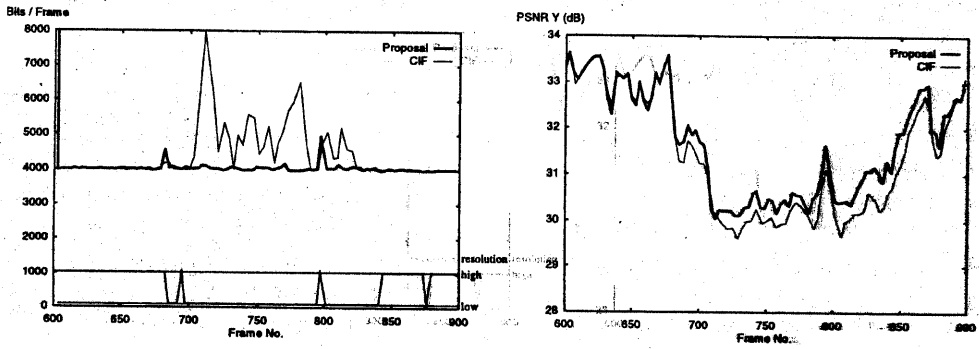


図9 "Mother & Daughter"の発生情報量, 解像度, 画質

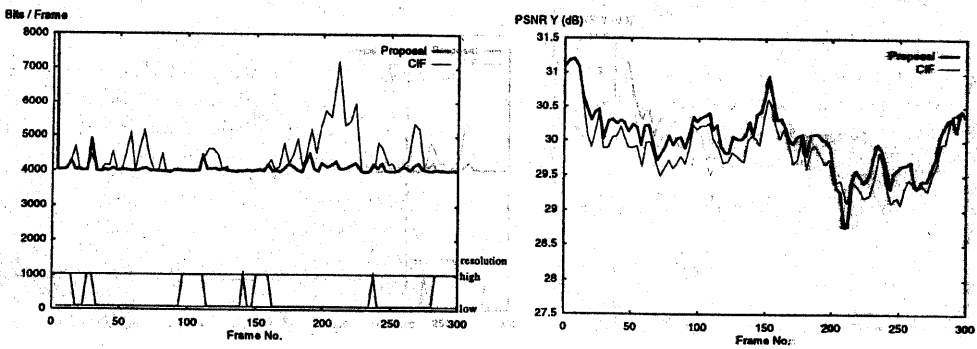


図10 "Silent Voice"の発生情報量, 解像度, 画質