

修士論文

ウェーブレットの帯域間相関を利用した
高精度動き推定に基づいた高能率動画像符号化

HIGH-EFFICIENCY VIDEO CODING BASED ON HIGH-PRECISION
MOTION ESTIMATION USING WAVELET'S SUBBAND
CORRELATION

2005年1月31日提出

指導教官 相田 仁 教授

東京大学大学院 新領域創成科学研究科 基盤情報学専攻

47-36310 佐藤 大将



目次

| | |
|--|----|
| 概要 | 1 |
| 第 1 章 序論 | 2 |
| 1.1 はじめに | 2 |
| 1.2 本論文における前提 | 3 |
| 1.3 本論文の構成 | 4 |
| 第 2 章 背景 | 5 |
| 2.1 ウェーブレット解析 | 6 |
| 2.2 ウェーブレット領域における動き推定・補償 | 13 |
| 2.2.1 空間領域における動き推定・補償 | 13 |
| 2.2.2 DWT 領域における動き推定・補償 | 14 |
| 2.2.3 ODWT 領域における動き推定・補償 | 15 |
| 2.2.4 比較実験 | 16 |
| 2.3 ODWT 参照フレームからの動き推定・補償 | 22 |
| 2.3.1 ウェーブレットブロック動き推定 | 22 |
| 2.3.2 Band-by-band 動き推定 | 24 |
| 2.3.3 比較実験 | 25 |
| 2.4 Multiresolution Motion Estimation (MRME) | 29 |
| 2.5 本研究の位置付け | 29 |
| 第 3 章 2 段階離散ウェーブレット変換 | 30 |
| 3.1 静止画像に対するウェーブレット変換の最適分解レベル | 31 |
| 3.2 動き推定・補償時の最適分解レベル | 34 |
| 3.3 2 段階離散ウェーブレット変換 | 37 |
| 3.4 実験 | 38 |
| 3.5 結論 | 41 |
| 第 4 章 動きベクトルの帯域間相関を利用した 2 段階 Band-by-band 動き推定 | 42 |
| 4.1 動きベクトルの帯域間冗長性 | 43 |
| 4.2 アルゴリズム | 43 |
| 4.3 Multiresolution Motion Estimation との違い | 44 |
| 4.4 メリット | 44 |
| 4.5 デメリット | 44 |

| | | |
|--------------|--------------------------------|-----------|
| 4.6 | 実験 | 45 |
| 4.7 | 結論 | 50 |
| 第 5 章 | LH1 バンドの係数値を用いた動的探索範囲設定 | 51 |
| 5.1 | 仮説 | 52 |
| 5.2 | 検証 | 52 |
| 5.3 | アルゴリズム | 54 |
| 5.4 | 実験 | 55 |
| 5.5 | 結論 | 59 |
| 第 6 章 | 結論 | 60 |
| 6.1 | まとめ | 60 |
| 6.2 | 今後の課題 | 61 |
| 6.2.1 | 動的探索範囲設定における適応的閾値設定 | 61 |
| 6.2.2 | 3 段階動き推定 | 61 |
| | 謝辞 | 62 |
| | 参考文献 | 63 |
| | 発表文献 | 66 |

目 次

| | | |
|------|---|----|
| 2.1 | フーリエ変換 特定の時刻の周波数を知ることができない。 | 6 |
| 2.2 | 短時間フーリエ変換 時間幅が固定となってしまう。 | 7 |
| 2.3 | ウェーブレット解析 高周波成分は短時間、低周波成分は長時間観測する。 | 7 |
| 2.4 | 1次元 DWT LPF と HPF とダウンサンプリングで構成できる。 | 11 |
| 2.5 | 2次元 DWT 画像のような二次元信号に対しては1次元 DWT を垂直・水平方向にそれぞれ行う。 | 12 |
| 2.6 | 2次元 DWT 実際の画像に適用するとこのようになる。LL 成分に信号が偏っていることがわかる。 | 12 |
| 2.7 | 空間領域における動き推定・補償 | 13 |
| 2.8 | DWT 領域における動き推定・補償 | 14 |
| 2.9 | Overcomplete Discrete Wavelet Transform によるウェーブレット分解 | 15 |
| 2.10 | ODWT 領域における動き推定・補償 | 15 |
| 2.11 | "football" に対する各領域動き補償の PSNR 比較 | 16 |
| 2.12 | "susie" に対する各領域動き補償の PSNR 比較 | 17 |
| 2.13 | "football" に対する各領域動き補償の主観的画質比較 (3Mbps、ブロックサイズ 16x16、サーチレンジ ± 16) | 18 |
| 2.14 | "football" に対する各領域動き補償の動きベクトルのビットレート比較 | 20 |
| 2.15 | "football" に対する各領域動き補償の平均処理時間比較 | 21 |
| 2.16 | ウェーブレットブロック動き推定 [19] (各サブバンドごとに独立に動き推定を行う。ブロックサイズは $B \times 2^{-l} (l = 1, 2, \dots, L)$) | 22 |
| 2.17 | Band-by-band 動き推定 [19] (各サブバンドごとに独立に動き推定を行う。ブロックサイズは $B \times 2^{-l} (l = 1, 2, \dots, L)$) | 24 |
| 2.18 | "football" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の PSNR 比較 | 25 |
| 2.19 | "susie" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の PSNR 比較 | 26 |
| 2.20 | "football" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の動きベクトルビットレート比較 | 27 |
| 2.21 | "football" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の平均処理時間比較 | 28 |
| 3.1 | "Lenna 512x512 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性 | 31 |
| 3.2 | "football#0 720x480 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性 | 32 |
| 3.3 | "susie#0 720x480 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性 | 32 |
| 3.4 | "mobile#0 720x576 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性 | 33 |

| | | |
|------|--|----|
| 3.5 | "football"シーケンスに対する分解レベル2・3における Band-by-band 動き推定の PSNR | 35 |
| 3.6 | "susie"シーケンスに対する分解レベル2・3における Band-by-band 動き推定の PSNR . . | 35 |
| 3.7 | "foreman"シーケンスに対する分解レベル2・3における Band-by-band 動き推定の PSNR | 36 |
| 3.8 | 2段階 DWT を用いた CODEC のブロック図 | 37 |
| 3.9 | "football"シーケンスに対する2段階 DWT の PSNR | 38 |
| 3.10 | "susie"シーケンスに対する2段階 DWT の PSNR | 39 |
| 3.11 | "foreman"シーケンスに対する2段階 DWT の PSNR | 39 |
| 3.12 | "football"シーケンスに対する2段階 DWT の平均処理時間 | 40 |
| 4.1 | Band-by-band 動き推定による動きベクトル | 43 |
| 4.2 | "football"に対する2段階 Band-by-band 動き推定の PSNR | 46 |
| 4.3 | "susie"に対する2段階 Band-by-band 動き推定の PSNR | 47 |
| 4.4 | "foreman"に対する2段階 Band-by-band 動き推定の PSNR | 48 |
| 4.5 | "football"に対する2段階 Band-by-band 動き推定の動きベクトルビットレート | 49 |
| 4.6 | "football"に対する2段階 Band-by-band 動き推定の平均処理時間 | 50 |
| 5.1 | "football"16 番フレームの DWT 係数 (左) と動きベクトル (右) | 53 |
| 5.2 | "susie"46 番フレームの DWT 係数 (左) と動きベクトル (右) | 53 |
| 5.3 | "foreman"1 番フレームの DWT 係数 (左) と動きベクトル (右) | 53 |
| 5.4 | "football"に対する動的探索範囲2段階 Band-by-band 動き推定の PSNR | 56 |
| 5.5 | "football"に対する動的探索範囲2段階 Band-by-band 動き推定の平均処理時間 | 57 |
| 5.6 | "susie"に対する動的探索範囲2段階 Band-by-band 動き推定の PSNR | 58 |
| 5.7 | "susie"に対する動的探索範囲2段階 Band-by-band 動き推定の平均処理時間 | 59 |

表 目 次

| | | |
|-----|---|----|
| 2.1 | 動き補償領域比較実験における平均 PSNR | 17 |
| 2.2 | ウェーブレットブロック動き推定と Band-by-band 動き推定の比較実験における平均 PSNR | 26 |
| 3.1 | 動き推定・補償時の最適分解レベル評価実験における平均 PSNR | 36 |
| 3.2 | 2 段階 DWT 評価実験における平均 PSNR (BB:Band-by-band, WB:Wavelet-block) | 40 |
| 4.1 | 2 段階 Band-by-band 動き推定評価実験における平均 PSNR (BB:Band-by-band, 2-BB:2 step BB, SR:Search Range) | 45 |
| 5.1 | 動的探索範囲評価実験における”football”シーケンスの平均 PSNR (2-BB:2 step Band-by-band, 2-BB:2 step BB, SR:Search Range) | 56 |
| 5.2 | 動的探索範囲評価実験における”susie”シーケンスの平均 PSNR (2-BB:2 step Band-by-band, 2-BB:2 step BB, SR:Search Range) | 58 |

概要

近年のネットワーク・蓄積メディアの多様化により、一つのビットストリームから必要な分だけを復号することにより様々な解像度・画質・フレームレートの動画を得ることができるスケーラブル動画像符号化が必要になってきている。本論文ではさらに高画質・高速を目指すための3つの手法を提案する。第一に、2段階離散ウェーブレット変換により静止画と動画のウェーブレット分解レベルのミスマッチを解決し、画質を向上させた。第二に、動きベクトルの帯域間相関を用いた2段階 Band-by-band 動き推定により動きの大きい動画に対して計算量を変えずに画質を向上させた。第三に、LH1 バンドの係数値を用いた動的探索範囲設定により画質をほとんど落とすことなく第二手法を高速化した。以上3手法をあわせて用いることにより、0.5～0.7dB 程度画質を向上させ、同時に全体の処理時間を46～49%に削減した。

ABSTRACT

Networks and storage media have been diversified, so scalable video coding is needed, which can get a variety of videos by decoding bits needed for a resolution, image quality and frame rate you want from one bitstream. In this paper, we propose three methods to get high image quality and high speed. First, "2 step Discrete Wavelet Transform" has solved a problem of a decomposition level mismatch between still images and movies, and it has improved image quality. Second, "2 step band-by-band motion estimation using subband correlation of motion vectors" has also improved image quality containing large motions without changing computation amount. Third, "Adaptive search range allocation using LH1-band coefficients" makes the second method without losing image quality. By using all the three methods at the same time, we can get PSNR improvement of 0.5 to 0.7 dB and reduce entire processing time to 46 to 49% simultaneously.

第1章 序論

1.1 はじめに

動画は写真を連続的に表示するものであるためそのままではデータ量が膨大である。このため動画の空間冗長性・時間冗長性を削減するいわゆる動画圧縮が何十年も研究されてきた。そしてこれまでの集大成とも言えるべきものが MPEG-4 AVC (Advanced Video Coding) / H.264 (以下、H.264) である。将来 H.264 より高圧縮な CODEC はなかなか生まれてこないだろう。またブロードバンド・蓄積メディア目覚ましい進歩からその必要もない。

そして近年、動画圧縮には拡張性が必要とされて来ている。ネットワークは光ファイバーから携帯電話網まで、蓄積メディアは Blu-ray ディスクから SD カードまでそれぞれ多様化している。通常、H.264 ではこれらにあわせて何度もエンコードを繰り返されなければならない。一方、現在標準化中である MPEG SVC(Scalable Video Coding, 以下、SVC) では一度エンコードすればそのビットストリームから必要な分だけ復号することにより様々な解像度・画質・フレームレートの動画を得ることができる。この性質をマルチキャストに利用し様々な環境にある受信側端末に一つのストリームを配信することにより無駄なトラフィックを防ぐこともできる。このような拡張性を実現する一つの手段がウェーブレット符号化である。ちなみに MPEG-2・H.264 にもこのような拡張性はあるがその性能はウェーブレット符号化に劣る。

離散ウェーブレット変換は信号を低周波成分と高周波成分に分解する。画像は水平・垂直方向に変換され 4 つのサブバンドに分解される。画像の場合水平・垂直両方向の低周波帯 (以下、LL バンド) に信号が集まり、縦・横それぞれ半分の解像度の画像となる。これを LL バンドに再帰的に行う。こうして現信号の主要成分を局所化し、かつ多重の解像度を得る。また、ウェーブレット係数をエンベデッド符号化し、画質の拡張性を得る。このようにウェーブレット符号化は様々な拡張性を持つ。

ウェーブレットのこれらの静止画像における拡張性を失うことなく動画像を符号化する研究が何年も行われている。本研究はこの分野に属し、特に動き推定・補償の点から高画質・高速を目指す。

1.2 本論文における前提

パラメータ設定

解像度は ITU-R 601 720x480(SDTV) を基本とする。これまでストリーミング等で広く用いられてきたのは SIF(352x288(NTSC), 352x240(PAL)) であるがネットワークの広帯域化・蓄積メディアの大容量化により SDTV が基本となるだろう。

ビットレートは 6Mbps を基本とする。今後蓄積メディアのスタンダードとなるであろう DVD 片面 2 層 8.5GB に 2 時間の動画、英・和の音声およびその他字幕などのテキスト情報を保存することを想定して 6Mbps とした。

フレームレートは 30fps・プログレッシブとする。実験に用いた動画のうち、“football”、“susie”(SDTV) はインタレースであるがこれをプログレッシブとみなした。

ブロックサイズは 16x16 とする。H.264 などでは可変ブロックサイズを採用しているが本論文の目的は動き推定の高精度化であるため、MPEG-2 と同じ 16x16 で固定とする。また、動きベクトルの符号化法において特に工夫していないため SDTV 解像度において Band-by-band 動き推定 (後述) を 8x8 のブロックサイズで行うと動きベクトルのレートだけで 6Mbps を超えてしまう。

サーチレンジは ± 16 を基本とする。SDTV 動画を基本として実験を行っているためサーチレンジは比較的大きくなくてはならない。

動き推定・補償は順方向予測のみを対象とする。

実装

実装は適応算術符号化のみ Amir Said 氏の Fast AC [37] を利用し、それ以外は全て筆者が行った。

評価方法

客観的評価の指標として PSNR(Peak Signal to Noise Ratio) を用いる。しかも本論文では輝度値 Y のみによる Y-PSNR で評価を行う。U と V を含めた PSNR に比べて色誤差を考慮に入れていない分、主観画質に多少の個人差がでるが人は色変化には鈍感で明るさの変化には敏感であるのでそれほど問題にはならない。

そもそも PSNR による評価は妥当かという議論もある。PSNR は画像全体の平均値なので、壁や人間の頬など変化の少ない部分が多いのっぺりした画像ほど PSNR が稼げる。逆に言えば、そのような画像ではエッジや細部などの高周波をどんなに正確に表現しても PSNR はそれほど上がらない。

PSNR にはこのような問題もあるが本論文では客観評価の一指標として用いることにする。

1.3 本論文の構成

本論文は6つの章から構成される。以下に各章の構成を示す。

第1章 研究の動機および本論文の構成について述べる。

第2章 ウェーブレットに基づいた動画像符号化、特に動き推定について研究動向をまとめる。

第3章 静止画像に対するウェーブレット変換の最適分解レベルと動画に対する動き推定・補償時の最適分解レベルをそれぞれ求めこのミスマッチを解決する2段階離散ウェーブレット変換を提案し、その構成図・実験結果を示す。

第4章 文献 [19] においてフルサーチであった Band-by-band 動き推定において、動きベクトルの帯域間相関を用いて2段階動き推定を提案する。これにより動きの大きい動画の動き推定においてサーチレンジを大きくすることなく、より精度の高い動きベクトルを求める。

第5章 前章において提案した2段階動き推定の高速化手法として LH1 バンドの係数値を用いた動的探索範囲手法を提案する。

第6章 本論文のまとめと今後の課題を述べる。

第2章 背景

MPEG では動き推定・補償を行ったあとに DCT を適用していた。DCT は位置情報を保存しない変換であり (フーリエ変換の欠点でもある)、変換後の係数はもはや原画像の原型を留めていない。そのため DCT の後に動き推定・補償を行うことができない。一方、ウェーブレット変換は位置情報を保存する変換であるため変換後も (サブバンド分割されているが) 画像である。そのため動き推定・補償を変換前か変換後に行うかの 2 通りの選択がある。

本章ではまず 2.1 節においてウェーブレット解析の基礎をまとめる。次に 2.2 節においてウェーブレット領域における動き推定・補償のメリットを述べ、2.3 節においてその関連研究をまとめる。

2.1 ウェーブレット解析

DCT(離散コサイン変換) はフーリエ変換の一種であるが、多くの信号に対して信号の周波数成分は非常に重要なためにフーリエ解析は非常に有効なものである。

それでは、なぜウェーブレット解析のような他の手法を必要とするのだろうか？

フーリエ変換

フーリエ解析には重大な欠陥がある。これは、周波数領域への変換で、時間情報が失われることである。信号のフーリエ変換を調べてみると、特別な事象が生じる時刻を知ることができない。(逆変換で元の信号が得られるので、正確に言えば時間情報は失われていないが、 $F(\cdot)$ の式だけではわからない。)

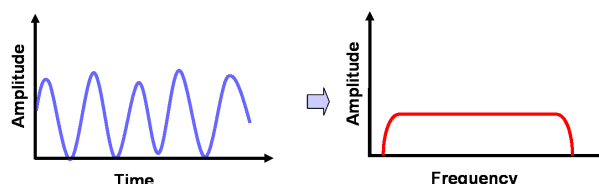


図 2.1: フーリエ変換 特定の時刻の周波数を知ることができない。

信号が、ある時間全体に渡って、あまり変化しない場合 (このことを 定常 という)、この欠点はあまり問題にはならない。しかし、ほとんどの興味ある信号は、かなりの非定常性を含んでいる。これらの特性が信号の最も重要な部分になることが多く、そのためフーリエ解析はこのような現象を調べるには適していない。(図 2.1)[4]

短時間フーリエ変換

この欠点を修正する 1 つの方法として、Dennis Gabor(1946) は、信号の小さな部分のみを解析するフーリエ変換 (信号のウィンドウ化と呼ばれる手法) を採用した。Gabor の方法は、いわゆる、短時間フーリエ (Short-Time Fourier Transform)(STFT) と呼ばれており、信号を時間と周波数の 2 次元関数に射影するものである。

STFT は、信号の時間ベ - スの見方と周波数ベ - スの見方の間の一種の解決策ではある。これは、ある信号イベントがいつ生じ、どの周波数であるかの 2 つについての情報を与える。しかし、この情報の精度はウィンドウの大きさに依存しているという制限がある。(図 2.2)

STFT により得られる時間情報と周波数情報は利用可能であるが、この欠点は時間ウィンドウに対してある特定の大きさを一度選択するとウィンドウはすべての周波数に対して同じように適用されることである。多くの信号はより柔軟性のあるアプロ - チ、すなわち、時間または周波数のどちらかをより正確に決定するためにウィンドウサイズを可変にできるアプロ - チを必要とする。[4]

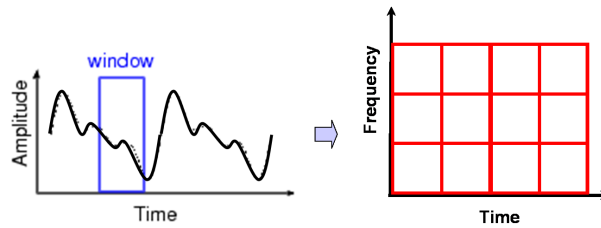


図 2.2: 短時間フーリエ変換 時間幅が固定となってしまう。

不確定性原理

時間-周波数解析において不確定性原理というものが存在し、窓関数 $w(t)$ の時間幅を $2\Delta_\omega$ 、そのフーリエ変換の帯域幅を $2\Delta_{\hat{\omega}}$ とすると、

$$\Delta_\omega * \Delta_{\hat{\omega}} \geq \frac{1}{2}$$

の関係が存在する。この式の両辺を 4 倍した式、

$$2\Delta_\omega * 2\Delta_{\hat{\omega}} \geq 2$$

であることから、時間-周波数平面において面積 2 の領域が時間と周波数を同時に詳しく知る限界である。
[5]

ウェーブレット解析

ウェーブレット解析は、つぎのような必然的に生じるステップ、すなわち、大きさを可変にできるウィンドウ操作手法を使う。ウェーブレット解析は、低い周波数情報をより詳細に得ようとする部分に対しては時間的に長い区間を取り、高い周波数情報をより詳細に得ようとする部分に対しては時間的に短い区間を考える。(図 2.3)[4]

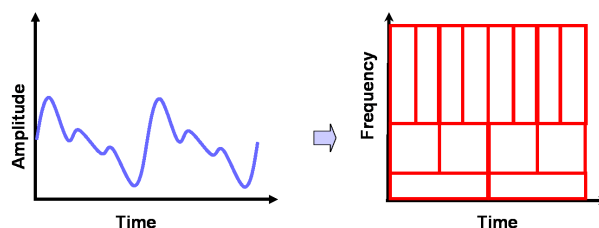


図 2.3: ウェーブレット解析 高周波成分は短時間、低周波成分は長時間観測する。

ウェーブレットとは

ところで、ウェーブレットとは何だろうか？

ウェーブレットとは、三角関数や対数関数のような特定の決まった関数ではなく、局在する波を表す様々な関数の総称である。語源は「波」"wave"と「小さな」"lets"の複合語であり、時間と周波数の小さなさざなみのことである。

観測された信号やパターン $f(x)$ に対してウェーブレットを用いてその信号 $f(x)$ の局所的様子を調べるためには、局所的な基本参照波 $\Psi(x)$ (これをマザーウェーブレットという) およびそれをスケール (伸縮)、トランスレート (平行移動、シフト) することで得られる参照波、

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right)$$

をいくつか用意して、これらとパターンとの関係を調べる。なお、基本参照波 $\Psi()$ は解析者により決定される。ここで、 b はトランスレート、 $a > 0$ はスケールのパラメータである $\frac{1}{\sqrt{a}}$ は正規化のための係数である。ただし $\Psi(x)$ は、

$$\int_{-\infty}^{\infty} \Psi(x) dx = 0$$

を満たす関数である。この式はウェーブレット条件と呼ばれ、ウェーブレットであるための基本条件であり、 $\Psi(x)$ が振動的であることを意味している。これからすぐわかるように、

$$\int_{-\infty}^{\infty} \Psi_{a,b}(x) dx = 0$$

が成り立つ。この式は、 $\Psi_{a,b}(x)$ が一種の高周波フィルタであることを意味している。この $\Psi_{a,b}(x)$ をマザーウェーブレットから産まれたウェーブレットという意味からドウターウェーブレットと呼ぶこともある。

連続ウェーブレット変換

連続ウェーブレット変換は、入力信号 $f(x)$ とウェーブレット $\psi_{a,b}(x)$ との内積で与えられる。すなわち、

$$\begin{aligned} W_{\psi} f(b, a) &= \langle f(x), \Psi_{a,b}(x) \rangle \\ &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(x) \overline{\Psi\left(\frac{x-b}{a}\right)} dx \end{aligned}$$

と表される。これらトランスレートパラメータ b とスケールパラメータ a を適切に決定することで、解析する波形中のこれと似たようなスケールの波形を、時間 (場所) 軸情報を失うことなく抽出することができるのである。定義式において積分範囲は $-\infty$ から ∞ となっているが、マザーウェーブレットはサポートコンパクト (範囲が限られている) なのでマザーウェーブレットの幅の範囲分だけ積分すればよい。ただし、マザーウェーブレットの幅はスケールパラメータ a によって変化する。[9]

離散ウェーブレット変換

連続ウェーブレットは、連続的な時間シフトと連続的なスケールをパラメータにもつ関数であるが、これらのパラメータが離散的な値をとるとき、離散ウェーブレットとよばれる。

$$a = \frac{1}{2^j}, b = \frac{k}{2^j}$$

とすると、離散ウェーブレットは、

$$\Psi_{j,k}(x) = 2^{\frac{j}{2}} \Psi(2^j x - k)$$

となり、(狭義の) 離散ウェーブレット変換は

$$d_k = 2^{\frac{j}{2}} \int_{-\infty}^{\infty} f(x) \overline{\Psi(2^j x - k)} dx$$

となる。シフトパラメータを動かすことにより時間あるいは場所を特定することができ、そこでスケールを変えると各レベルの高周波成分を取り出すことができる。スケールは2のべき乗であるので、 j が解像度レベルとなる。 j が小さいほど解像度は小さくなる。

また、離散ウェーブレットではマザーウェーブレットとする関数は基底関数となるものでなければならない。[9]

多重解像度解析の原理

では、解像度スケーラビリティを得るための多重解像度解析の原理を説明する。

一般に、ツースケール関係、

$$\phi(x) = \sum_{k \in Z} p_k * \phi(2x - k)$$

を満たす関数 ϕ をスケーリング関数という。数列 $\{p_k\}$ はツースケール数列と呼ばれ、数列 $\{p_k\}$ によってスケーリング関数 ϕ は決定される。 $\{\phi(x - k) | k \in Z\}$ は正規直交系とする。

ここでツースケール関係より

$$\phi(2^j x) = \sum_{k \in Z} p_k * \phi(2^{j+1} x - k)$$

が成り立つことから、

$$V_j = \text{Span}\{\phi(2^j x - k) | k \in Z\}$$

とすると

$$V_j \subset V_{j+1}$$

がいえる。すると、Hilbert の直交分解定理より V_{j+1} は

$$V_{j+1} = V_j \oplus W_j, V_j \perp W_j$$

のように一意に直交分解できる。 W_j は V_{j+1} における V_j の直交補空間とよばれ、

$$W_j = \text{Span}\{\psi(2^j x - k), k \in Z\}$$

のように、ウェーブレット関数 $\psi(x)$ を 2^j だけスケールし、それをシフトした関数系によって張られる。この分解の繰り返しを用いれば、原信号の空間 V_1 は、

$$\begin{aligned} V_1 &= V_0 \oplus W_0 \\ &= (V_{-1} \oplus W_{-1}) \oplus W_0 \\ &= \bigoplus_{j=-\infty}^0 W_j \end{aligned}$$

のように分解できる。そして $W_{j-1} \in V_j$ と直交関係 $V_j \perp W_j$ により直交関係 $W_{j-1} \perp W_j$ が得られるので、この分解は互いに直交するウェーブレット空間による直交分解を表している。このように信号の空間を次々と低周波成分の空間と高周波成分の空間に直交分解していく一連の操作を多重解像度解析と呼ぶ。

Mallat 変換

多重解像度解析の原理を用いて、実際に原信号を低周波数と高周波数の成分に分ける変換を Mallat 変換 (またはオクターブツリー分解または広義の離散ウェーブレット変換) という。Mallat 変換に関しては、ツースケール数列とウェーブレット数列のみを含んだ高速の分解公式と復元公式が知られている。まずツースケール数列とウェーブレット数列を定義する。

ツースケール関係は、

$$\phi(x) = \sum_{k \in \mathbb{Z}} p_k * \phi(2x - k)$$

であり、また、 $W_0 \oplus V_0 = V_1$ より、 $W_0 \subset V_1$ であるから、 $\psi \in V_1$ は

$$\psi(x) = \sum_{k \in \mathbb{Z}} q_k * \phi(2x - k)$$

と表され、数列 p_k をツースケール数列、数列 q_k をウェーブレット数列と呼ぶ。

次にこれらの数列を用いた Mallat 変換を簡単に示す。

V_j 、 W_j の任意の元 $f_j \in V_j$, $g_j \in W_j$ は

$$f_j(x) = \sum_{k \in \mathbb{Z}} c_k^{(j)} * 2^{\frac{j}{2}} \phi(2^j x - k)$$

$$g_j(x) = \sum_{k \in \mathbb{Z}} d_k^{(j)} * 2^{\frac{j}{2}} \psi(2^j x - k)$$

と表せる。ここで V_{j+1} は $V_{j+1} = V_j \oplus W_j$ と分解できることより、

$$f_{j+1}(x) = f_j(x) + g_j(x)$$

となるから、 $c_k^{(j+1)}$ から $c_k^{(j)}$ と $d_k^{(j)}$ が求められれば、低周波成分をさらなる低周波成分と高周波成分に分けられたことになる。これらは、

$$c_k^{(j)} = \frac{1}{2} \sum_l \overline{p_{l-2k}} c_l^{(j+1)}$$

$$d_k^{(j)} = \frac{1}{2} \sum_l \overline{q_{l-2k}} c_l^{(j+1)}$$

によって計算できることが示されている。これを Mallat 変換という。

まず、原画に Mallat 変換を適用すると、第 1 低周波成分と第 1 高周波成分が得られる。さらに、第 1 低周波成分を原画像とみなして分解公式を適用すると、第 2 低周波成分と第 2 高周波成分が得られる。このように次々に分解公式を適用すると、たった一つの低周波成分と、解像度の異なった高周波成分の集まりが得られる。もちろん復元公式を用いると原画像を復元できるので、分解によって得られたこのような低周波成分と高周波成分の集まりは原画像と同等である。[7]

これら分解および再構成アルゴリズムは観測パターンの圧縮符号化として知られているサブバンド符号化と関係している。すなわち、一般的な直交ウェーブレット関数はサブバンド符号で用いるフィルタバンクから構成できる。例えば、最も簡単な 2 チャンネルのフィルタバンクの場合、一連の観測パターンを入力すると、入力系列の標本化レートの半分のレートになった 2 チャンネルの系列を出力する。サブバンド符号化ではこれらの二つの系列を符号化し、復号化する復号化された 2 チャンネル系列を標本化レートが 2 倍のフィルタに通し、元の入力系列（観測パターン）を再生する。ここで用いるフィルタは低域、高域のフィルタである。フィルタのインパルス応答は、変換の定義から、基底関数とみなすことができるので、このようなフィルタバンクは入力系列を基底関数によって低域と高域に分解していることと等価である。出力した 2 チャンネルの系列のうち低域フィルタの出力に繰り返しフィルタバンクを用いると、周波数分解能は上がるが時間分解能は下がる。この状況はウェーブレット変換と同じであり、この操作自体がウェーブレット変換と等価である。[9] 図 2.4 はこの操作を表しており、0L は原信号、1L は第一階低周波成分、1H は第一階高周波成分、2L は第二階低周波成分、2H は第二階高周波成分を表す。

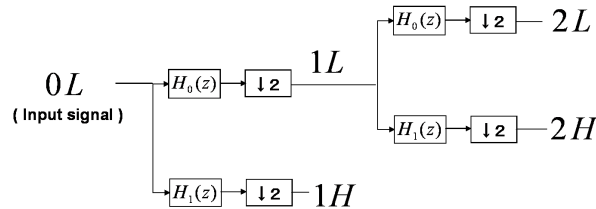


図 2.4: 1 次元 DWT LPF と HPF とダウンサンプリングで構成できる。

画像は2次元であるので2次元の離散ウェーブレット変換が必要となる。これは1次元の離散ウェーブレット変換を垂直方向・水平方向に行う。この結果原画像は、4つの周波数成分(サブバンド)に分解される。(図2.5)

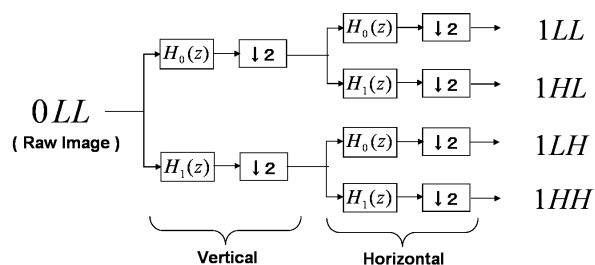


図 2.5: 2次元 DWT 画像のような二次元信号に対しては1次元 DWT を垂直・水平方向にそれぞれ行う。

実際の画像に適用すると図2.6のようになる。LL成分に主要な信号が偏っていることがわかる。

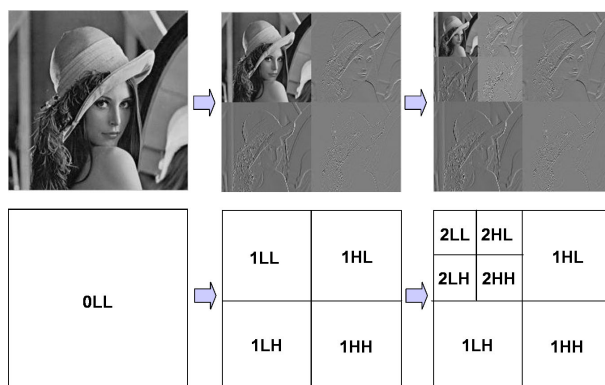


図 2.6: 2次元 DWT 実際の画像に適用するとこのようになる。LL成分に信号が偏っていることがわかる。

このように離散ウェーブレット変換により、さまざまな解像度の画像を容易に得ることができる。次節ではウェーブレットに基づいた動画像符号化について述べる。

2.2 ウェーブレット領域における動き推定・補償

前節で述べたようにウェーブレットは位置情報を保存する。そのため MPEG と異なり動画画像符号化における動き推定・補償を変換前に空間領域で行うか、変換後に変換領域で行うかの2通りの選択がある。本節ではそれぞれの実現法、利点・問題点を述べる。

2.2.1 空間領域における動き推定・補償

DWT を用いたエンコーダで第一に思いつくのは図 2.7 のようなものだろう。これは MPEG などのエンコーダの DCT を DWT に変えただけの構成である。ここで SPIHT(Set Partitioning in Hierarchical Trees)[18] はウェーブレット係数のエンコーダである。また、AAC(Adaptive Arithmetic Coding) は適応算術符号化を表す。

しかしながらこのエンコーダには問題がある。動き補償後の画像は不連続部が多いため、DWT を画像全体にかけるとこの不連続部が高周波になってしまう。高周波は SPIHT により量子化され劣化または失われる。このため不連続部付近で符号化効率が落ちてしまう。

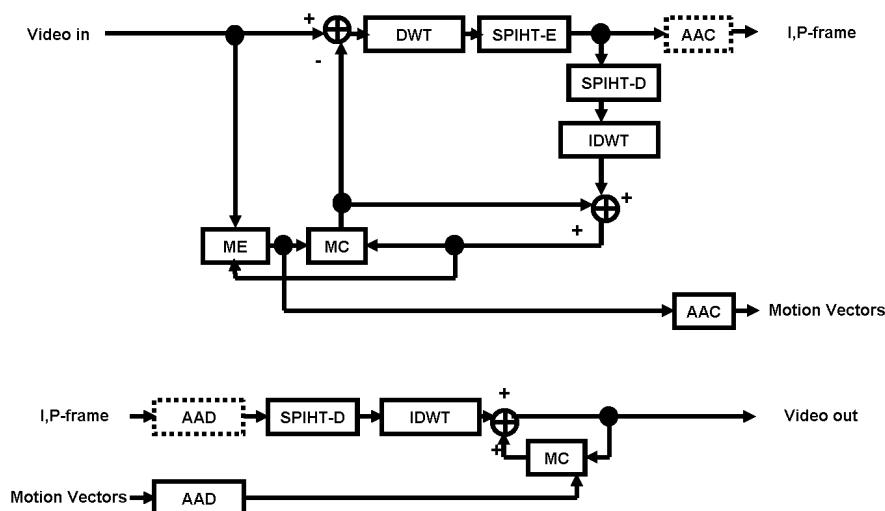


図 2.7: 空間領域における動き推定・補償

2.2.2 DWT 領域における動き推定・補償

そこでウェーブレット領域で動き推定・補償を行う。つまりエンコーダを図 2.8 のような構成とする。すると補償による不連続な部分が変換されることはない。またエンコーダに IDWT を必要としないというメリットもある。

しかしながら、ここでも問題がある。DWT は時変性 (*shift-variance*) を持つ。

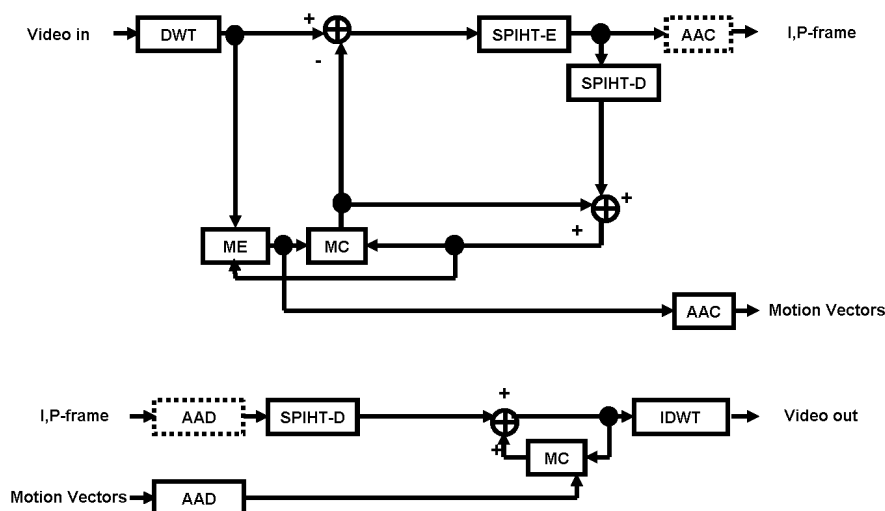


図 2.8: DWT 領域における動き推定・補償

離散ウェーブレット変換の時変性

システムの入力と出力間の特性が入力の加えられる時間によって変化するシステムを時変システムという。[10]

ここで、DWT は時変システムであることに注意しなければならない。すなわち、

$$WaveletCoefficients(x - dx) \neq DWT(InputSignal(x - dx))$$

これは DWT におけるダウンサンプリングが原因である。というのは、例えばダウンサンプリングにより、現画像の 1 ピクセルの動きは正しくサブバンド画像に伝わらないからである。低周波領域ではこの影響は少ないが、高周波領域 (エッジなど) ではかなり大きな影響を受ける。そのため、ウェーブレット領域における動き推定・補償には工夫が必要である。

2.2.3 ODWT 領域における動き推定・補償

DWT はダウンサンプリングのせいで時変システムとなってしまった。ではダウンサンプリングを止めよう、というのが Overcomplete Discrete Wavelet Transform (以下、ODWT) の基本的な考え方である。そのため、各サブバンド画像のサイズは現画像のサイズと同じである。(図 2.9) ODWT は別名、"Redundant DWT", "Undecimated DWT", "the algorithme a trous" ともいう。[22]

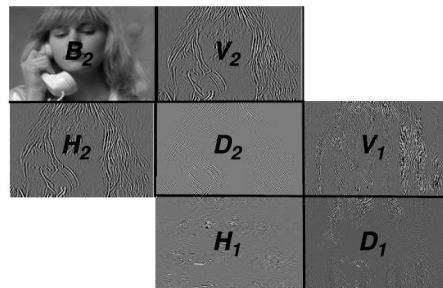


図 2.9: Overcomplete Discrete Wavelet Transform によるウェーブレット分解

ただし、全ての DWT を完全に ODWT に置き換えるわけではない。参照フレームのみに完全な係数が残っていればよいので、動き推定を行う直前に Inverse DWT・ODWT を行うのである。このエンコーダを図 2.10 に示す。(ちなみに ODWT の実装法として Low-band-shift 手法 [19] というものが有名であるが、あくまで実装法であり画質は ODWT の場合と変わらないのでここでは省略する。)

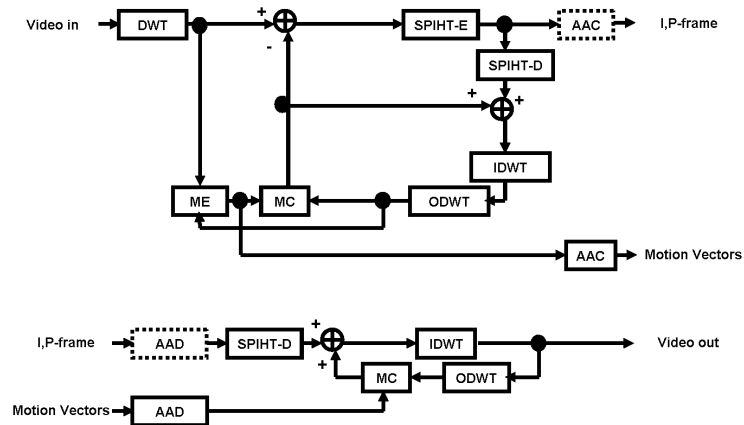


図 2.10: ODWT 領域における動き推定・補償

2.2.4 比較実験

各領域において動き推定・補償を行う CODEC を図 2.7・2.8・2.10・のように定義し、”CIPR ITU601 football 720x480”・”CIPR ITU601 susie 720x480”シーケンスを用い比較実験を行った。YCbCr4:2:0 変換を行い、ビットレートは 6Mbps (I フレームのみ 2.0bpp で符号化、それ以外は誤差画像と動きベクトルのビットレート)。またブロックサイズは 16x16、サーチレンジは ± 16 、ウェーブレット分解レベルは文献 [19] と同じ 3 にした。

実験結果を、”football”・”susie”シーケンスに対する PSNR をそれぞれ図 2.11・2.12、平均 PSNR を表 2.1 に示す。また”football”の動きベクトルビットレート・平均処理時間を 2.14・2.15 に示す。

さらに主観評価のため、実際の画像を図 2.13 に示す。ただし主観評価用のみ 3Mbps でエンコードしている。

画質

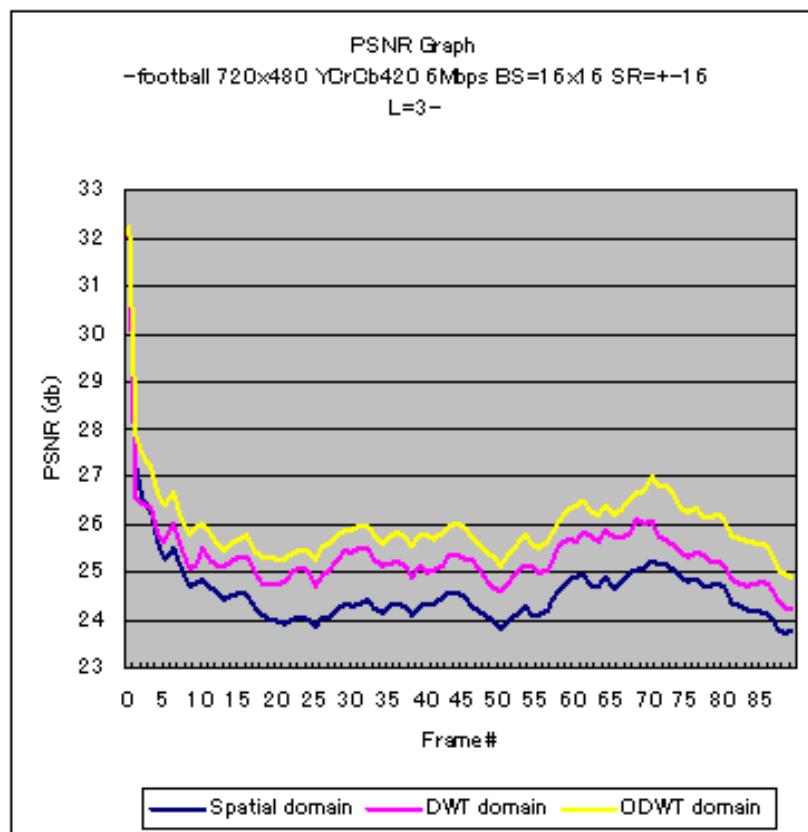


図 2.11: ”football”に対する各領域動き補償の PSNR 比較

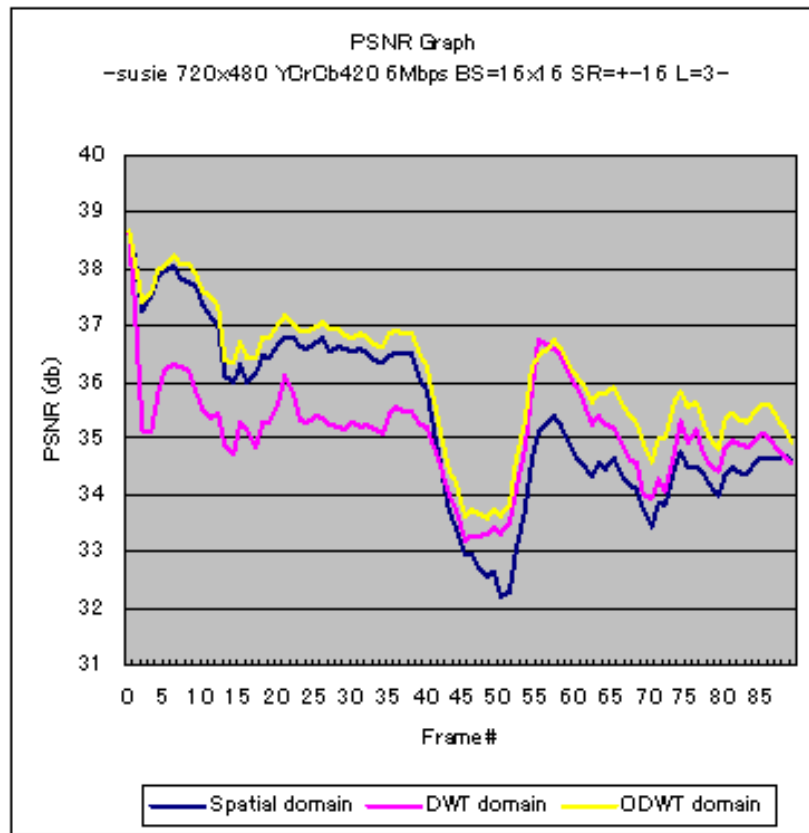


図 2.12: "susie" に対する各領域動き補償の PSNR 比較

表 2.1: 動き補償領域比較実験における平均 PSNR

| Sequence Name | Spatial domain | DWT domain | ODWT domain |
|---------------|----------------|------------|-------------|
| football | 24.64 | 25.35 | 25.99 |
| susie | 35.38 | 35.12 | 36.09 |



Original



Spatial domain (25.80db)



DWT domain (24.75db)



ODWT domain (26.36db)

図 2.13: "football" に対する各領域動き補償の主観的画質比較 (3Mbps、ブロックサイズ 16x16、サーチレンジ ± 16)

図 2.13 からわかるように、空間領域動き補償ではブロックノイズが出ており、DWT 領域動き補償では shift-variance の影響で画質が劣化している。これに対し ODWT 領域ではブロックノイズもなく shift-variance の影響もないので最も高画質となる。この原因を考察する。

”football”シーケンスは全体的に動きが大きい。動きベクトル場がなめらかでないため動き補償画像は不連続部が多くなる。この不連続部は差分画像にも受け継がれ、変換・量子化により失われる。このため再生画像はブロック境界が目立つ。”susie”シーケンスは全体的に動きが少ないがおよそ 45 番フレームくらいから 60 番フレームくらいにかけて動きが比較的大きい。この部分で先程と同様のことが起こってブロックノイズが目立ち、画質が劣化する。一方 DWT・ODWT 領域では変換を先に行うためこのような問題は起こらず、ブロックノイズは発生しない。

また、DWT 領域動き補償は shift-variance 問題が起こり画質が劣化する。これは主観的には画質が波打っているように見える。分解レベル 3 の場合、DWT 領域の 1 ピクセルの動きは空間領域では 8 ピクセルに対応するため、動き推定ミスが起こるとかなり遠くのブロックを補償してしまう。この結果、画質はつながりの悪いものとなる。一方、ODWT 領域では完全な係数が残っているため ODWT 領域の 1 ピクセルシフトは空間領域でも 1 ピクセルシフトでありこのような shift-variance 問題は起こらない。

これらの原因から ODWT 領域動き補償が数値上でも主観的にも最も高画質となる。

動きベクトルのビットレート

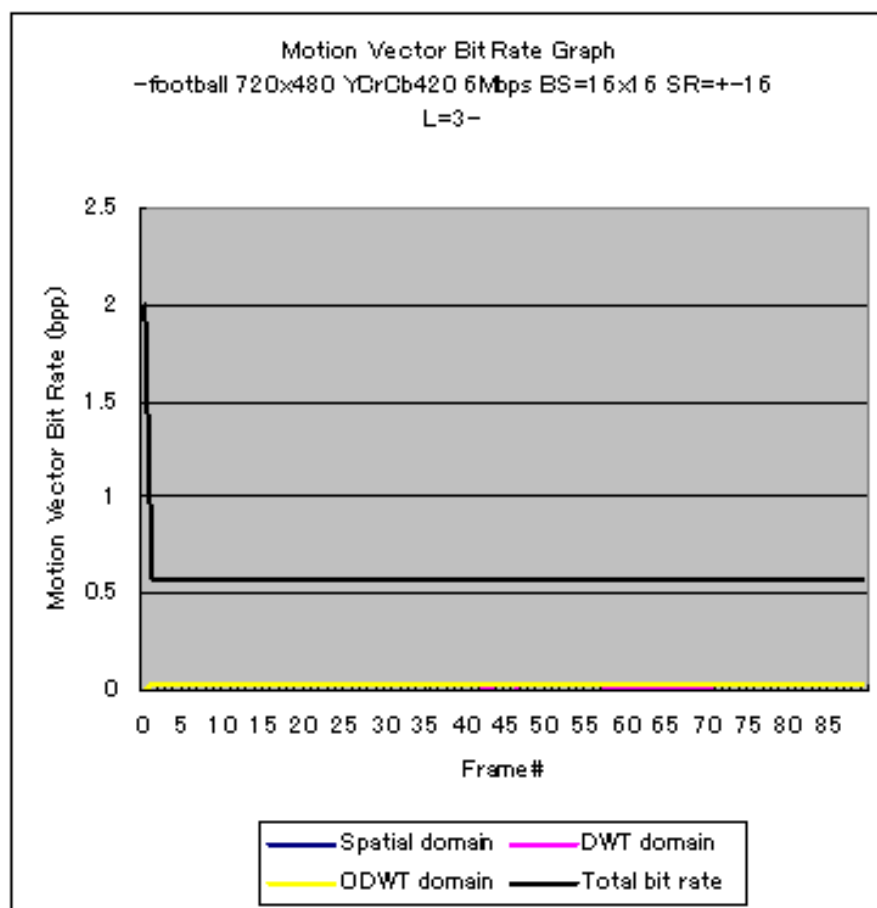


図 2.14: "football" に対する各領域動き補償の動きベクトルのビットレート比較

三者とも動きベクトルの数は変わらないので動きベクトルのビットレートもほとんど変わらない。ただし DWT 領域動き補償はサーチレンジが小さく動きベクトルの種類が少ないためエントロピー符号化によりやや動きベクトルビットレートは下がる。

平均処理時間

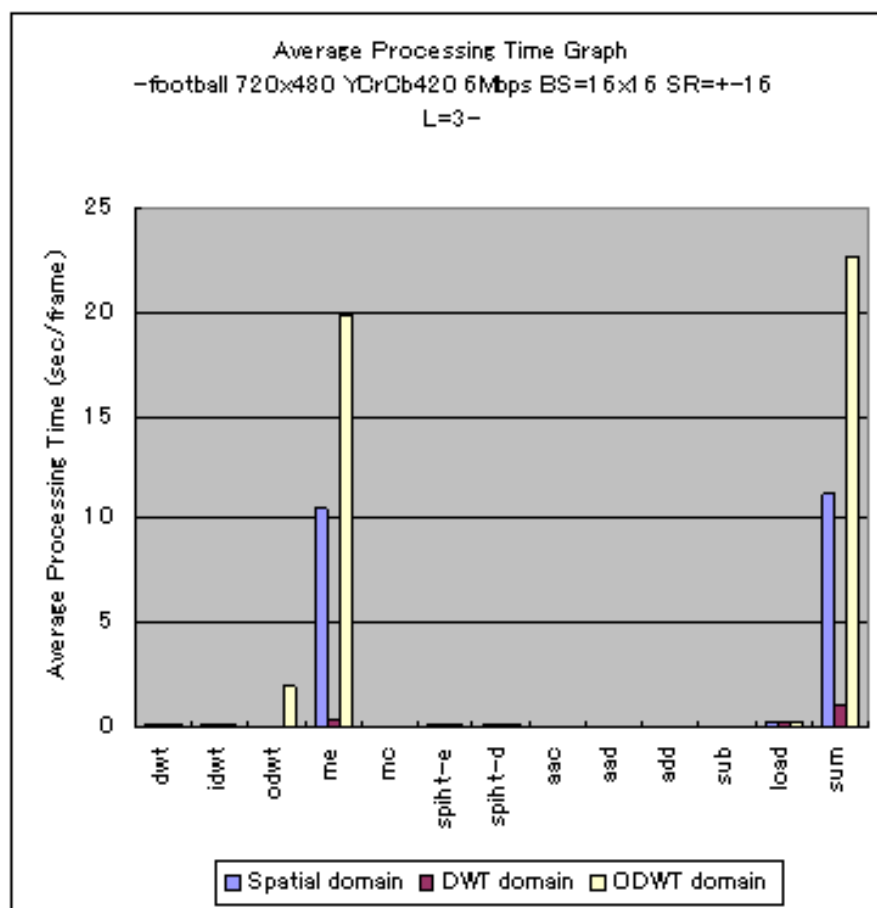


図 2.15: "football" に対する各領域動き補償の平均処理時間比較

ODWT 領域動き補償は ODWT を行う必要があり、この計算時間が約 2 秒とかなり大きい。ODWT はデコードにも必要なため再生すらできない。しかし今回の実験の ODWT の遅さは筆者の実装の仕方が原因である。高速化の手法はいくつかあるが、JPEG2000 など使われているリフティングアルゴリズム [1] を用いる等すればもっと速くなるだろう。

また、DWT 領域動き推定がとても早く見えるかもしれない。これは空間領域でのサーチレンジ ± 16 は分解レベル 3 の DWT 領域では ± 2 に該当するからである。

2.3 ODWT 参照フレームからの動き推定・補償

ODWT フレームから DWT フレームを予測する方法は Park と Kim により 2 通りが提案されている。[19]

2.3.1 ウェーブレットブロック動き推定

動画画像符号化におけるフレーム間の冗長性を減らすため、動体は主に平行移動の動きをするという仮定が一般になされている。このため、ブロックマッチングアルゴリズムが空間領域において画像ブロックの動きベクトルを見積もるのに一般に適用される。

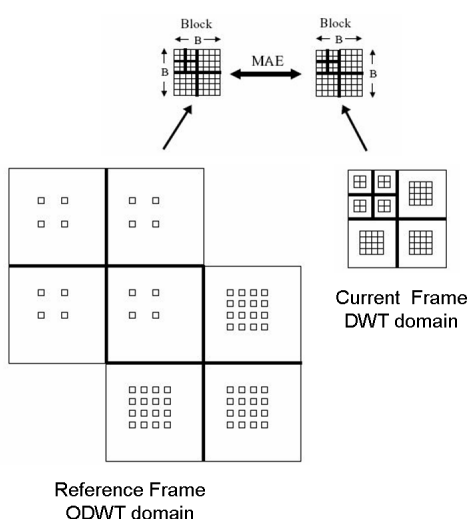


図 2.16: ウェーブレットブロック動き推定 [19] (各サブバンドごとに独立に動き推定を行う。ブロックサイズは $B \times 2^{-l} (l = 1, 2, \dots, L)$)

2次元画像の DWT 係数は位置情報と周波数帯域を持っているため、DWT 係数は空間領域でひとつの画像ブロックになるように空間位置によってブロックデータをグループ化することができる。LL バンドに根を持つ各ブロックの DWT 係数は図 2.16 のように再構成される。ウェーブレットブロックの目的はウェーブレット係数とそれが画像において空間的にあらわすものを直接的に連想させることである。全てのスケール・位置における関連した係数は各ブロックに含まれる。例えばレベル 3 のウェーブレット変換を画像に用いた場合、 16×16 のウェーブレットブロックはレベル 1 の LH, LH, HH バンドの 8×8 ブロックと、レベル 2 の LH, LH, HH バンドの 4×4 ブロックと、LL, LH, LH, HH バンドの 2×2 ブロックから構成される。

参照フレーム内のサーチウィンドウ W におけるウェーブレットブロックは現在のウェーブレットブロックと比べられ、最も適合する参照ウェーブレットブロックが選ばれる。例えば入力画像がレベル 3 で分解された場合、入力画像は計 10 個のサブバンドに分解される。すなわち、3 つのレベル 1 のサブバンドと 3 つのレベル 2 のサブバンド、そして 4 つのレベル 3 のサブバンドである。変位ベクトルを (dx, dy) とおくと、ウェーブレットブロックの MAE (Mean Absolute Error) は次のように計算される。

$$\begin{aligned}
MAE(x, y, dx, dy) &= \frac{2^{-L}}{B^2} \sum_{i=0}^{B/2^L-1} \sum_{j=0}^{B/2^L-1} \left| LL_{cur}^{(L)}\left(\frac{x}{2^L} + i, \frac{y}{2^L} + j\right) - LL_{ref}^{(L)}(x + 2^L i + dx, y + 2^L j + dy) \right| \\
&+ \frac{1}{B^2} \sum_{l=1}^L 2^{-l} \sum_{i=0}^{B/2^l-1} \sum_{j=0}^{B/2^l-1} \left\{ \left| LH_{cur}^{(l)}\left(\frac{x}{2^l} + i, \frac{y}{2^l} + j\right) - LH_{ref}^{(l)}(x + 2^l i + dx, y + 2^l j + dy) \right| \right. \\
&+ \left| HL_{cur}^{(l)}\left(\frac{x}{2^l} + i, \frac{y}{2^l} + j\right) - HL_{ref}^{(l)}(x + 2^l i + dx, y + 2^l j + dy) \right| \\
&+ \left. \left| HH_{cur}^{(l)}\left(\frac{x}{2^l} + i, \frac{y}{2^l} + j\right) - HH_{ref}^{(l)}(x + 2^l i + dx, y + 2^l j + dy) \right| \right\}
\end{aligned}$$

上式を用いて最適な動きベクトル (dx_{min}, dy_{min}) は次式で与えられる。

$$(dx_{min}, dy_{min}) = \min_{(dx, dy) \in W} MAE(x, y, dx, dy)$$

ここで W は動き推定におけるサーチウィンドウである。

2.3.2 Band-by-band 動き推定

もうひとつは Band-by-band 動き推定と呼ばれる方法である。ウェーブレットに基づいた動き推定・補償の中で最も高画質になる手法である。これは変換領域と空間領域でブロックの適合する場所が必ずしも同じではない、という性質を利用したものである。

各サブバンドごとに SAD(Sum of Absolute Difference) により動き推定・補償を行い、それぞれのサブバンドの各ブロックに最適な動きベクトルを得る。(図 2.17 参照) 各サブバンドでブロックサイズは $B \times 2^{-l} (l = 1, 2, \dots, L)$ である。ここで B は空間領域でのブロックサイズを表す。ウェーブレットブロック動き推定と異なる点はひとつのウェーブレットブロックに対して複数の ($3L + 1$ 個、 L はウェーブレット分解レベル) 動きベクトルをそれぞれを独立に計算することである。

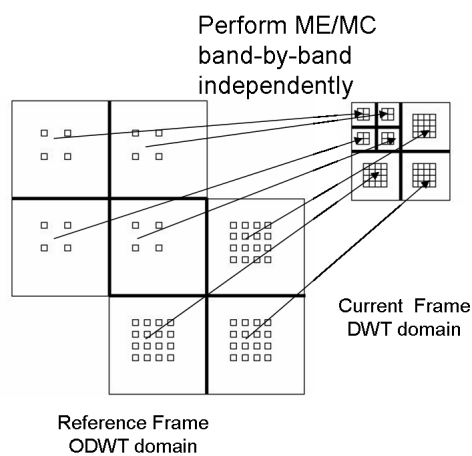


図 2.17: Band-by-band 動き推定 [19] (各サブバンドごとに独立に動き推定を行う。ブロックサイズは $B \times 2^{-l} (l = 1, 2, \dots, L)$)

動きベクトルの個数が増えてしまうという問題があるが、この手法は特に高ビットレートの動画などに有効である。高ビットレートの場合、動きベクトルのデータ量は多くても動き推定の精度が上がるため差分画像のデータ量が小さくなり総合的に高画質になるからである。

2.3.3 比較実験

両手法の CODEC を共に図 2.10 のように定義する。"CIPR ITU601 football 720x480"・"CIPR ITU601 susie 720x480"シーケンスを用い比較実験を行った。YCbCr4:2:0 変換を行い、ビットレートは 6Mbps (I フレームのみ 2.0bpp で符号化、それ以外は誤差画像と動きベクトルのビットレート)。またブロックサイズは 16x16、サーチレンジは ± 16 、ウェーブレット分解レベルは文献 [19] と同じ 3 にした。

実験結果を、"football"シーケンスに対する PSNR・動きベクトルビットレート・平均処理時間をそれぞれ図 2.18・2.20・2.21 に示す。また、"susie"シーケンスに対する PSNR を図 2.19 に示す。さらに両者の平均 PSNR を表 2.2 に示す。

画質

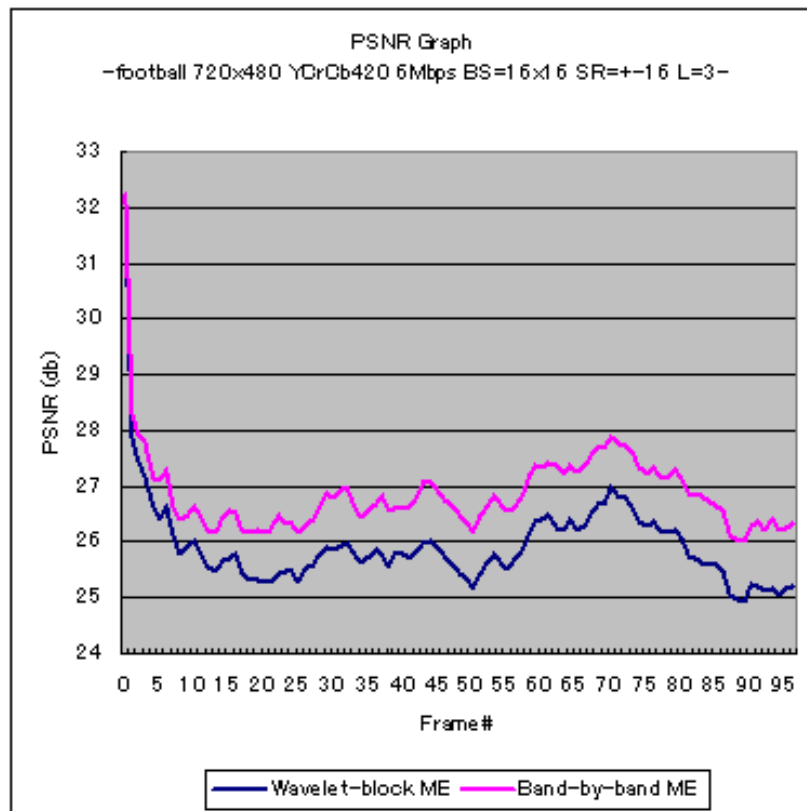


図 2.18: "football"に対するウェーブレットブロック動き推定と Band-by-band 動き推定の PSNR 比較

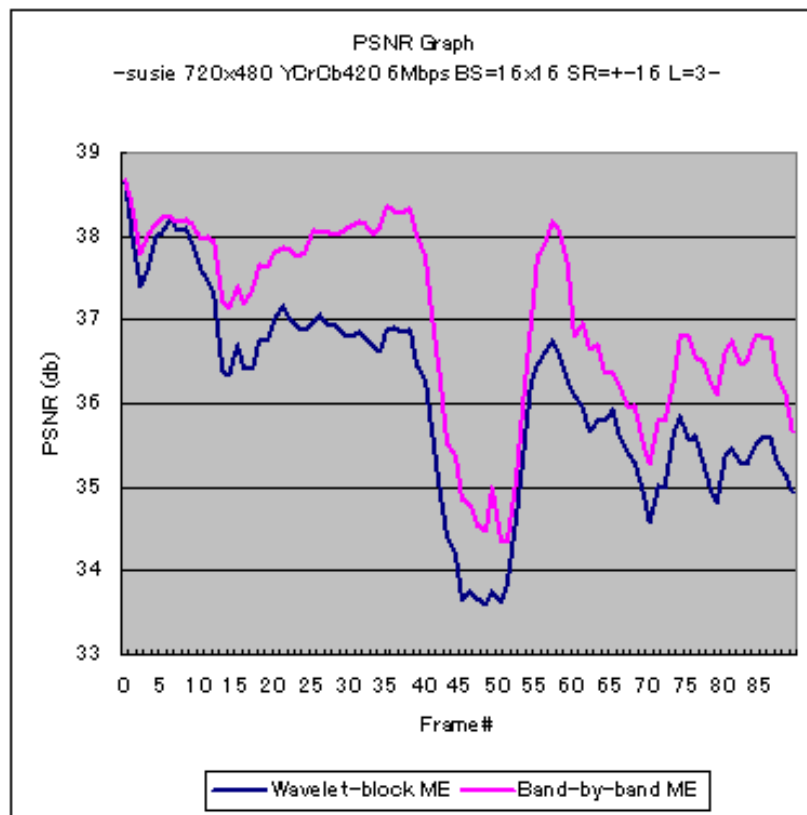


図 2.19: "susie" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の PSNR 比較

表 2.2: ウェーブレットブロック動き推定と Band-by-band 動き推定の比較実験における平均 PSNR

| Sequence Name | Wavelet-block ME | Band-by-band ME |
|---------------|------------------|-----------------|
| football | 25.93 | 26.86 |
| susie | 36.09 | 37.01 |

図 2.18・2.19・表 2.2 から両動画において Band-by-band 動き推定の方が高画質であることがわかる。これは Band-by-band 動き補償ではウェーブレットブロックを分解し各サブバンドで独立に動き補償を行っているため、ウェーブレットブロックの動き補償誤差がウェーブレットブロック動き補償よりも減っているからである。

動きベクトルのビットレート

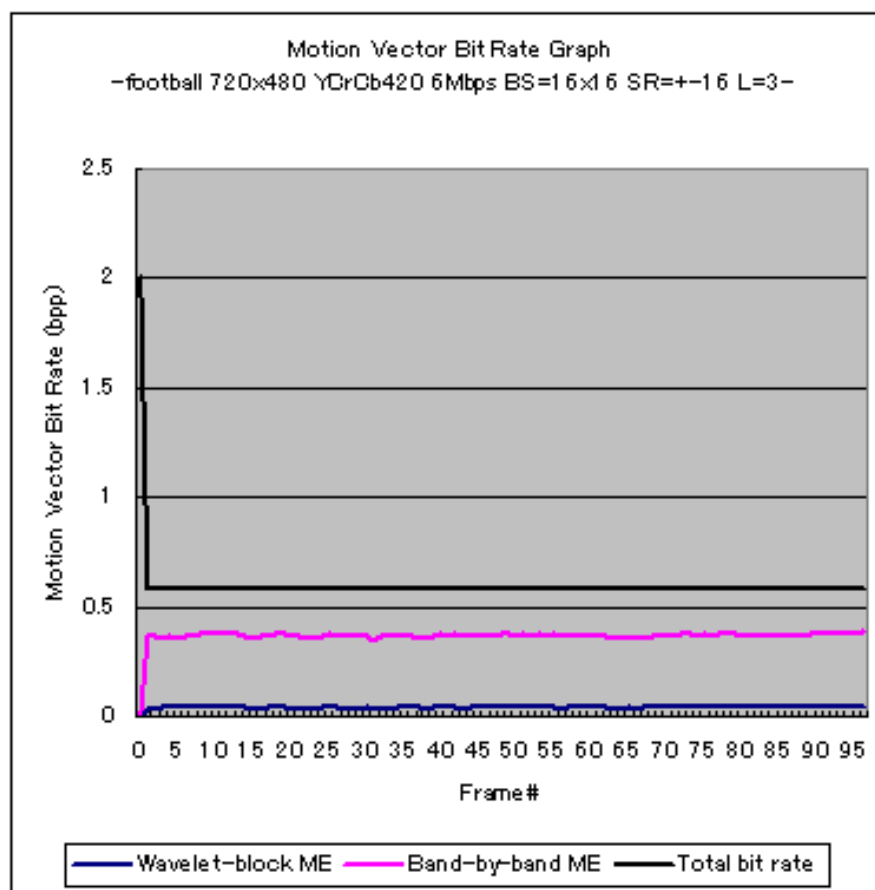


図 2.20: "football" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の動きベクトルビットレート比較

図 2.20 から Band-by-band 動き推定の動きベクトルビットレートがウェーブレットブロック動き推定に比べてかなり高いことがわかる。これは Band-by-band 動き推定はウェーブレットブロック動き推定に比べてこの場合（分解レベル 3 の場合）10 倍の数の動きベクトルを持つからである。そのため差分画像に割り当てられるビット量が少ない。しかし Band-by-band 動き推定は差分画像のエネルギーが小さいため、総合的には高画質となる。

平均処理時間

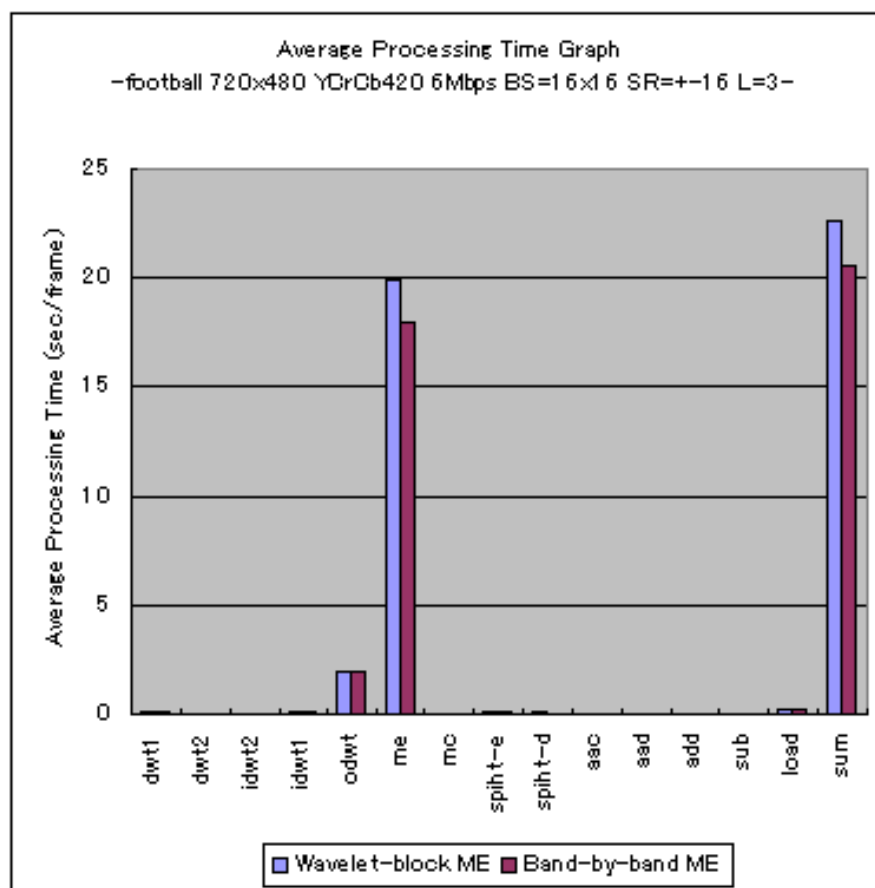


図 2.21: "football" に対するウェーブレットブロック動き推定と Band-by-band 動き推定の平均処理時間比較

動き推定にかかる時間は若干ではあるが Band-by-band 動き推定の方が速い。

2.4 Multiresolution Motion Estimation (MRME)

文献 [23]-[24] では Multiresolution Motion Estimation (MRME) が提案されている。DWT 領域において、まず LL バンドで動きベクトルを求め、他サブバンドではこれを拡大したものをサーチの初期位置として探索を行う。MRME は本論文第 4 章における提案手法と似ているが、MRME は DWT 領域で行われていること、LL バンド動きベクトルを拡大する点が異なる。また、shift-variance 問題にも対応していない。

また文献 [25] では、従来の MRME で coarse-to-fine 方式 (LL バンド動きベクトルが初期位置) を用いているのに対し、fine-to-coarse 方式を用いて画質を向上させている。まず空間領域で動きベクトルを求め、これを縮小したものを各サブバンドの初期位置とするものである。本論文の提案手法でも検討したがよい結果が得られなかったので省略した。またこちらも DWT の shift-variance 問題に対応していない。

2.5 本研究の位置付け

本研究はウェーブレットに基づいた動画像符号化方式分野において、Band-by-band 動き推定を高画質、さらに高速にすることを目指す。

第3章 2段階離散ウェーブレット変換

本章では2段階離散ウェーブレット変換を提案する。静止画像に対するDWTは分解レベル5が最適であるが動き補償においては分解レベル2が適していることが実験からわかった。この問題を解決すべく、DWTを2段階に分けた。

まず3.1節において静止画像に対する最適分解レベルを求める。次に3.2節においてBand-by-band動き推定に対する動き推定・補償時の最適分解レベルを求める。そしてこのミスマッチを解決する2段階DWTの構成図を3.3節に示し、3.4節において実験を行う。最後に結論を述べる。

3.1 静止画像に対するウェーブレット変換の最適分解レベル

離散ウェーブレット変換はLLバンドに対して再帰的に変換を行っていくのであるが、何回変換を行うのか、言い換えれば分解レベルをいくつにするのが最適であるかの検討が必要である。これはSPIHTなどのウェーブレット係数符号化における4分木の数に関係する。分解レベルが低いと4分木の数が増え、木ごとの相関がまだ残っている可能性がある。逆に分解レベルが高すぎると相関のない4分木まで一つの木とみなしてしまい無駄が多くなる。そこで実験を行い最適分解レベルを得る。ちなみにJPEG2000ではデフォルト分解レベルを5としている。JPEG2000ではウェーブレット係数の符号化にEBCOT(Embedded Block Coding with Optimal Truncation)を用いており多少状況は異なる。

標準画像”Lenna 512x512 RGB”、”football#0 720x480 RGB”、”susie#0 720x480 RGB”、”mobile#0 720x576 RGB”に対し実験を行った。結果をそれぞれ図 3.1 ~ 3.4 に示す。

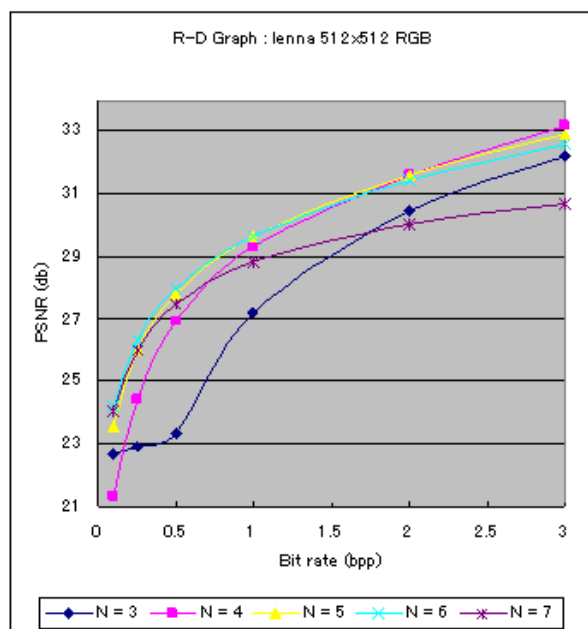


図 3.1: ”Lenna 512x512 RGB”に対するウェーブレット分解レベルごとのレート・ひずみ特性

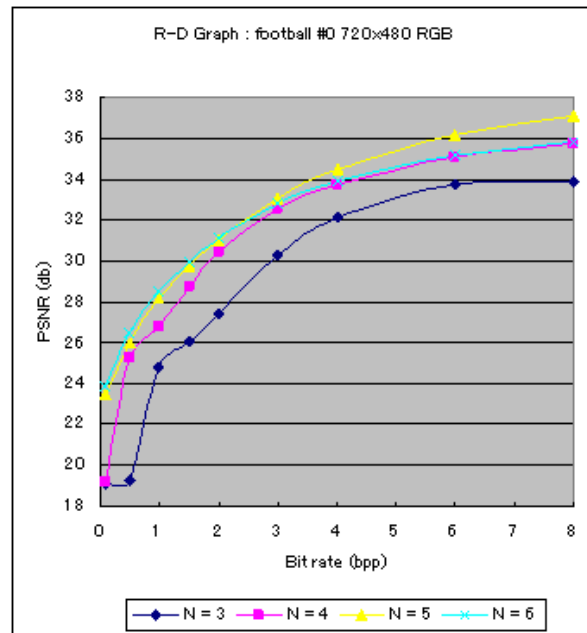


図 3.2: "football#0 720x480 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性

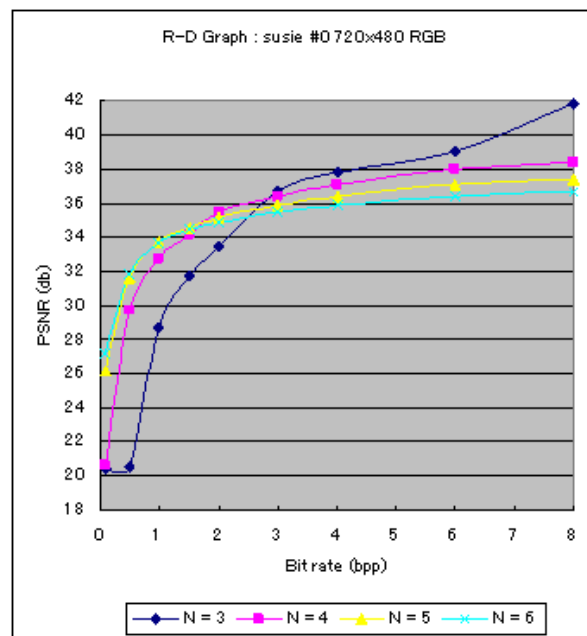


図 3.3: "susie#0 720x480 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性

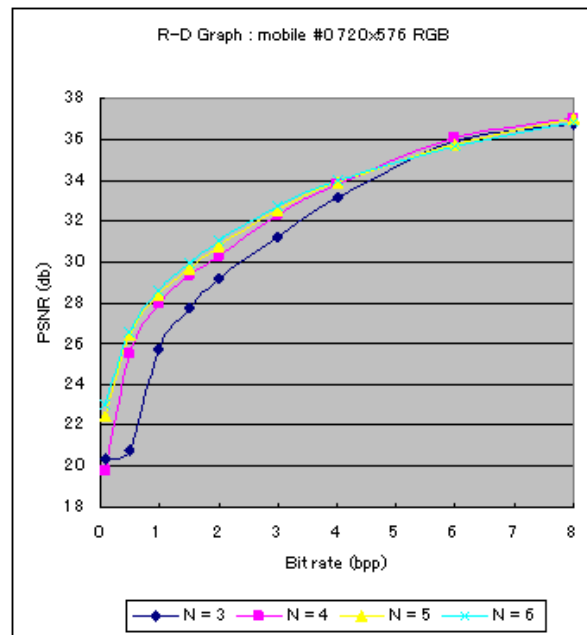


図 3.4: "mobile#0 720x576 RGB" に対するウェーブレット分解レベルごとのレート・ひずみ特性

全体的に分解レベル4以上からほぼ同じ符号化効率になってくるが、分解レベル4は低ビットレートにおいて符号化効率が低い。よって静止画像においては分解レベルは5が最適であると判断する。

3.2 動き推定・補償時の最適分解レベル

動き補償においても分解レベル5が最適であるとは言えない。例えばウェーブレットブロック動き推定の場合、LLバンドブロックサイズは最低でも1x1であり、分解レベル5・YCbCr4:2:0における最小ウェーブレットブロックサイズは64x64となる。ブロックサイズが大きすぎてブロックの適合率がかなり低くなってしまう。

そこでまずウェーブレットブロック動き推定と Band-by-band 動き推定の両者に対して以下の条件でウェーブレット分解レベルの最適化を行った。

両手法の CODEC を共に図 2.10 とする。”CIPR ITU601 football 720x480”・”CIPR ITU601 susie 720x480”・”CIPR SIF foreman 352x288”シーケンスを用い実験を行った。YCbCr4:2:0 変換を行い、ビットレートは6Mbps (Iフレームのみ2.0bppで符号化、それ以外は誤差画像と動きベクトルのビットレート)、またブロックサイズは16x16、サーチレンジは±16とした。

ここで、ウェーブレット分解レベルは3までしかないことに注意されたい。YCbCr4:2:0のためCb・Crコンポーネントのブロックサイズは8x8となるからである。また分解レベル1ではIフレームの画質が落ちすぎてしまうため分解レベルが最適であるはずはないので除外した。これらの理由から分解レベル2と3について比較を行った。

結果のPSNRを図3.5・3.6・3.7に、平均PSNRを表3.1示す。

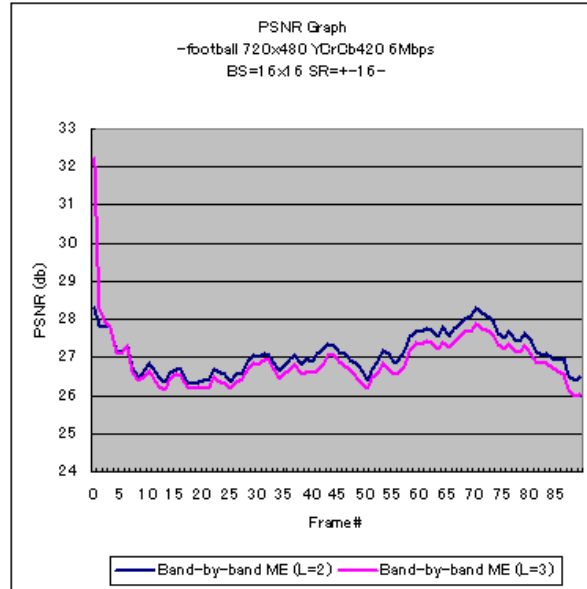


図 3.5: "football"シーケンスに対する分解レベル 2・3 における Band-by-band 動き推定の PSNR

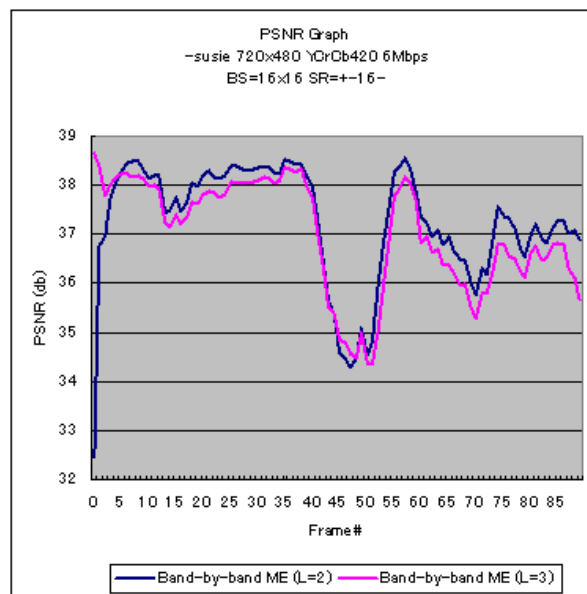


図 3.6: "susie"シーケンスに対する分解レベル 2・3 における Band-by-band 動き推定の PSNR

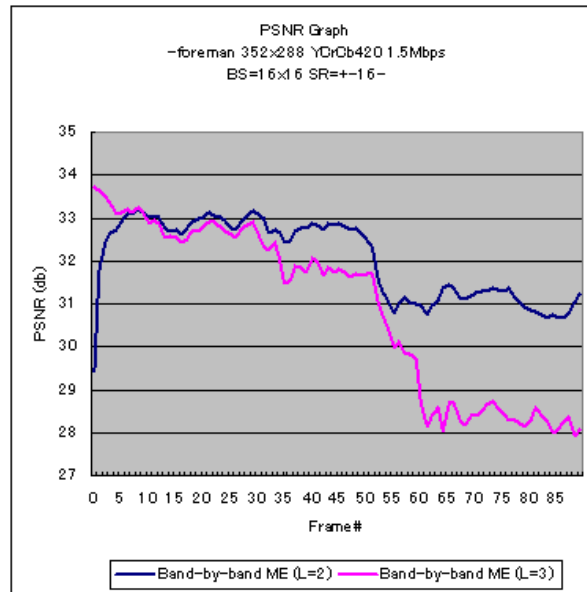


図 3.7: "foreman" シーケンスに対する分解レベル 2・3 における Band-by-band 動き推定の PSNR

表 3.1: 動き推定・補償時の最適分解レベル評価実験における平均 PSNR

| Sequence Name | Band-by-band ME (L=2) | Band-by-band ME (L=3) |
|---------------|-----------------------|-----------------------|
| football | 27.08 | 26.86 |
| susie | 37.24 | 37.01 |
| foreman | 32.03 | 30.90 |

実験結果から動き推定・補償時の最適分解レベルは 2 であると判断する。また、分解レベル 2 だと ODWT が高速になるのでエンコード・デコード両方にとって大きなメリットにもなる。

3.3 2 段階離散ウェーブレット変換

前節までから「静止画像は分解レベル 5 が最適であるのに対し動き補償は分解レベル 2 が最適である」という観測結果が得られた。そこでこの問題を解決するため図 3.8 のような構成を提案する。

まず原画像に対して DWT を分解レベル 1 ~ 2 まで行い、動き補償後にさらに DWT 分解 3 ~ 5 までを行う。

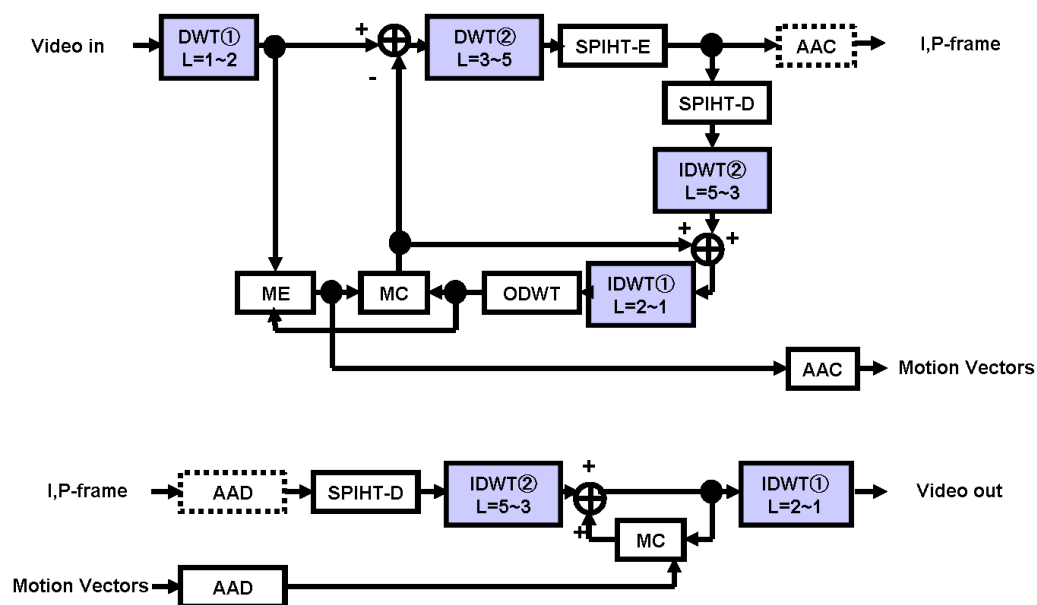


図 3.8: 2 段階 DWT を用いた CODEC のブロック図

3.4 実験

ウェーブレットブロック動き推定と Band-by-band 動き推定の両者に対して以下の条件で 2 段階 DWT の評価実験を行った。1 段階の場合の CODEC は共に図 2.10 とし、2 段階の場合の CODEC は図 3.8 とする。”CIPR ITU601 football 720x480”・”CIPR ITU601 susie 720x480”・”CIPR SIF foreman 352x288”シーケンスを用い実験を行った。YCbCr4:2:0 変換を行い、ビットレートは”football”と”susie”は 6Mbps、”foreman”は 1.5Mbps (I フレームのみ 2.0bpp で符号化、それ以外は誤差画像と動きベクトルのビットレート) とした。またブロックサイズは 16x16、サーチレンジは ± 16 とした。

この結果の PSNR を図 3.9・3.10・3.11、平均 PSNR を表 3.2、”football”における平均処理時間を図 3.12 に示す。

ここで 2 段階の最終レベルは 5 とする予定だったが、実装上の問題から 4 までしか分解できなかった。例えば 720x480 の解像度の場合、特に色差コンポーネントは分解レベル 3 を行った時点で LL バンドの解像度は 45x30 となる。信号幅が奇数となるため筆者の実装ではこれ以上分解できない。色差コンポーネントまでレベル 5 で分解しようとするとな元の解像度は 64x64 の倍数とならなければならない、720x480 の場合、解像度を 768x512 に拡張しなければならない。これでは画像が大きすぎるため今回は分解レベル 5 はあきらめ分解レベル 4 とした。これなら解像度は 32x32 の倍数となればよく、720x480 解像度は 736x480 と少しの拡張ですむ。今回の拡張は画像の右側に 0 を付加するゼロパディングとした。多少画質に影響があるとしても 3.1 節の実験で分解レベル 4 でも十分な符号化効率を得られていたので 2 段階 DWT を十分に評価できるだろう。

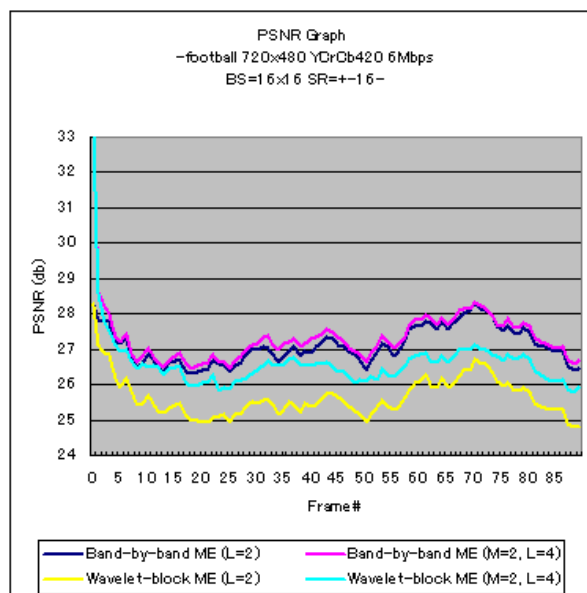


図 3.9: ”football”シーケンスに対する 2 段階 DWT の PSNR

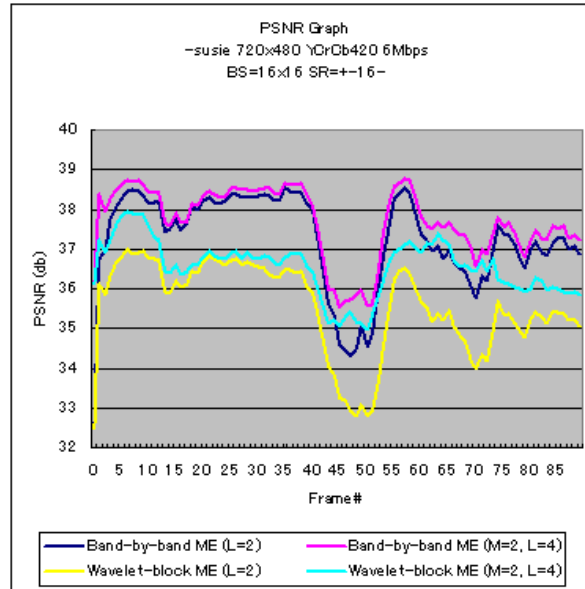


図 3.10: "susie"シーケンスに対する 2 段階 DWT の PSNR

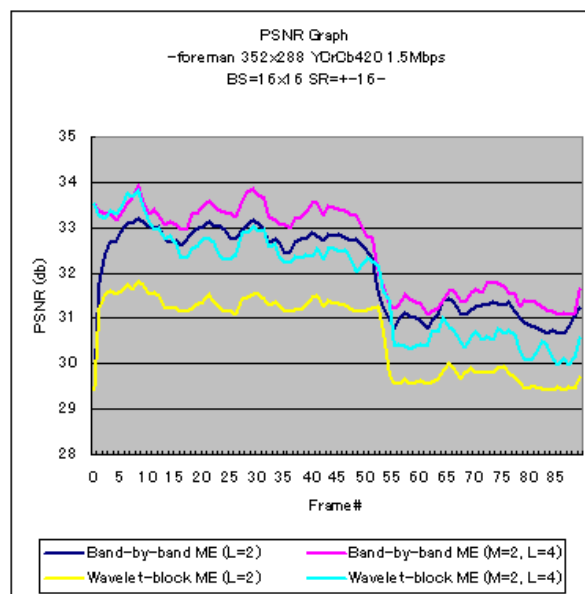


図 3.11: "foreman"シーケンスに対する 2 段階 DWT の PSNR

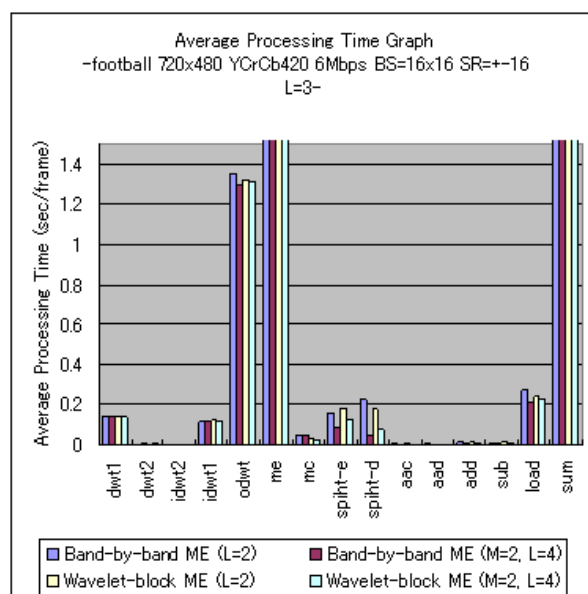


図 3.12: "football"シーケンスに対する 2 段階 DWT の平均処理時間

表 3.2: 2 段階 DWT 評価実験における平均 PSNR (BB:Band-by-band, WB:Wavelet-block)

| Sequence Name | BB ME | BB ME (2 step DWT) | WB ME | WB ME (2 step DWT) |
|---------------|-------|--------------------|-------|--------------------|
| football | 27.12 | 27.35 | 25.65 | 26.61 |
| susie | 37.24 | 37.69 | 35.54 | 36.54 |
| foreman | 32.03 | 32.52 | 30.65 | 31.81 |

実験結果から全ての動画に対してウェーブレット動き推定では約 1dB、Band-by-band 動き推定では約 0.2~0.4dB の PSNR の向上が得られた。両動き推定の PSNR 向上に差があるのはおそらくウェーブレット動き推定の方が誤差画像のエネルギーが大きいため、レベル 2 とレベル 4 の SPIHT の結果に差が出やすいからだろう。Band-by-band 動き推定は誤差画像のエネルギーが小さいため多少の効果しかない。おそらく両者において、高速だが精度が低い探索を行った場合またはサーチレンジ自体が狭いときに誤差画像のエネルギーが大きくなるため 2 段階 DWT は大きく性能を発揮するだろう。

3.5 結論

本章ではまず静止画像に対する最適ウェーブレット分解レベルを求めた。次に動画に対する動き補償時の最適分解レベルを求めた。これらの観測結果から静止画像は分解レベル5が最適であるのに対し動き補償は分解レベル2が最適であるという結論に達した。そこでこれらの問題を同時に解決する2段階DWTによるCODECの構成を提案した。実験結果はウェーブレットブロック動き推定・Band-by-band動き推定共に画質を向上させることができた。

本論文では以下、2段階DWTを行うという前提で議論を進め、実験を行う。

第4章 動きベクトルの帯域間相関を利用した2段階 Band-by-band 動き推定

文献 [19] で提案されたウェーブレットブロック動き推定と Band-by-band 動き推定は共にフルサーチである。ただのフルサーチは大きい動きに弱い。なぜなら計算量が2乗のオーダーで増えるためサーチウィンドウをあまり大きくできないからである。サーチウィンドウが小さいと大きな動きをカバーすることができない。

本章では大きい動きに対応するため、動きベクトルの帯域間相関を利用した2段階 Band-by-band 動き推定を提案する。

4.1 動きベクトルの帯域間冗長性



図 4.1: Band-by-band 動き推定による動きベクトル

Band-by-band 動き推定を分解レベル L で行った場合、動きベクトルはサブバンドの数と同じで $(3L+1)$ 組できる。各サブバンドの動きベクトル間に相関はあるだろうか？”football”シーケンス、 $L = 2$ の場合の各サブバンドの動きベクトルの例を図 4.1 に示す。一見するとばらばらのようにも見えるが、動きの大きい部分の動きベクトルは全体的に似ている。この動きベクトルの帯域間相関を用いる。

4.2 アルゴリズム

LL バンドは現信号が最も集中するサブバンドであり、この動きベクトルが最も実際の動きをよく表す。そこで次のような LL バンド動きベクトルを基本ベクトルとする 2 段階 Band-by-band 動き推定を提案する。

1. LL バンドでフルサーチにより動きベクトル $mv_{LL}(x, y)$ を得る。
2. 他サブバンドでは LL バンドで得られた動きベクトルの終点をサーチの中心としてフルサーチを行う。

$$\begin{aligned} mv_{the\ others}(x) &= mv_{LL}(x) + dx \\ mv_{the\ others}(y) &= mv_{LL}(y) + dy \end{aligned}$$

4.3 Multiresolution Motion Estimation との違い

Multiresolution Motion Estimation (MRME)[23]-[24] は DWT 領域で行われていた。本手法は ODWT 領域で行うため他サブバンドの探索の初期位置の計算方法が異なる。また、DWT の shift-variance 問題が起こらないため高画質である。

4.4 メリット

この手法のメリットは二つある。

一つは計算量を変えずに大きい動きに対応できること。LL バンド以外は大きい動きを追いかけてサーチウィンドウを設定しているのにより誤差の少ないブロックを見つけることができる。これにより誤差画像のエネルギーが小さくなり、画質が向上する。

もう一つは動きベクトルのビットレートが増えずにすむこと。一般に大きい動きをカバーするためにはその分大きなサーチウィンドウが必要になるがそうすると動きベクトルの種類が増えてしまいエントロピー符号化した際にデータ量が増えてしまう。提案手法ではサーチウィンドウサイズは変わっていないのでこのようなことはない。また、MPEG でも本手法と同じような方法は可能であるが、MPEG の場合基本動きベクトルを差分動きベクトルとは別に保存しておく必要がある。本手法では LL バンドが基本ベクトルであるためこのようなことはない。

4.5 デメリット

一方、この手法には一つのデメリットがある。

SAD (絶対値誤差総和) によるブロックマッチングは必ずしも実際の動きを表すわけではない。このため、LL バンドで実際の動きとは異なった動きベクトルを得た場合、他サブバンドにおけるサーチの中心は適さない場所になりかねない。この問題は分解レベルが高いときにおこりやすい。というのは、例えば分解レベル 4・ブロックサイズ 16x16 の場合、LL バンドにおけるブロックサイズは 1x1、つまり画素であり本手法において基本となる LL バンド動きベクトルは動きではなく似た色を指すものとなる可能性が高いからである。

本手法と基本アイデアは同じである手法 [23]-[25] ではこれが問題となるが、しかし、前章で提案した 2 段階 DWT では動き推定を分解レベル 2 で行うのでこの問題は起こりにくい。これも 2 段階 DWT のメリットとなるだろう。

4.6 実験

CODEC は図 3.8 とする。"CIPR ITU601 football 720x480"・"CIPR ITU601 susie 720x480"・"CIPR SIF foreman 352x288"シーケンスを用い実験を行った。YCbCr4:2:0 変換を行い、ビットレートは 6Mbps (I フレームのみ 2.0bpp で符号化、それ以外は誤差画像と動きベクトル)。またブロックサイズは 16x16、サーチレンジは $\pm 16 \cdot \pm 8 \cdot \pm 4$ とした。

ここで 2 段階 DWT の最終レベルは 5 とする予定だったが、実装上の問題から 4 までしか分解できなかった (理由は 3.4 節で示した)。多少画質に影響があるとしても本手法を十分に評価できるだろう。

"football"・"susie"・"foreman"の平均 PSNR を表 4.1、PSNR をそれぞれ図 4.2・4.3・4.4、"football"の動きベクトルビットレート・平均処理時間を図 4.2・4.3 に示す。

表 4.1: 2 段階 Band-by-band 動き推定評価実験における平均 PSNR (BB:Band-by-band, 2-BB:2 step BB, SR:Search Range)

| Sequence Name | BB (SR:16) | 2-BB (SR:16) | BB (SR:8) | 2-BB (SR:8) | BB (SR:4) | 2-BB (SR:4) |
|---------------|------------|--------------|-----------|-------------|-----------|-------------|
| football | 27.30 | 27.41 | 26.92 | 27.21 | 26.32 | 26.58 |
| susie | 37.69 | 37.78 | 37.25 | 37.53 | 36.71 | 36.94 |
| foreman | 33.35 | 33.35 | 33.74 | 33.70 | 33.80 | 33.75 |

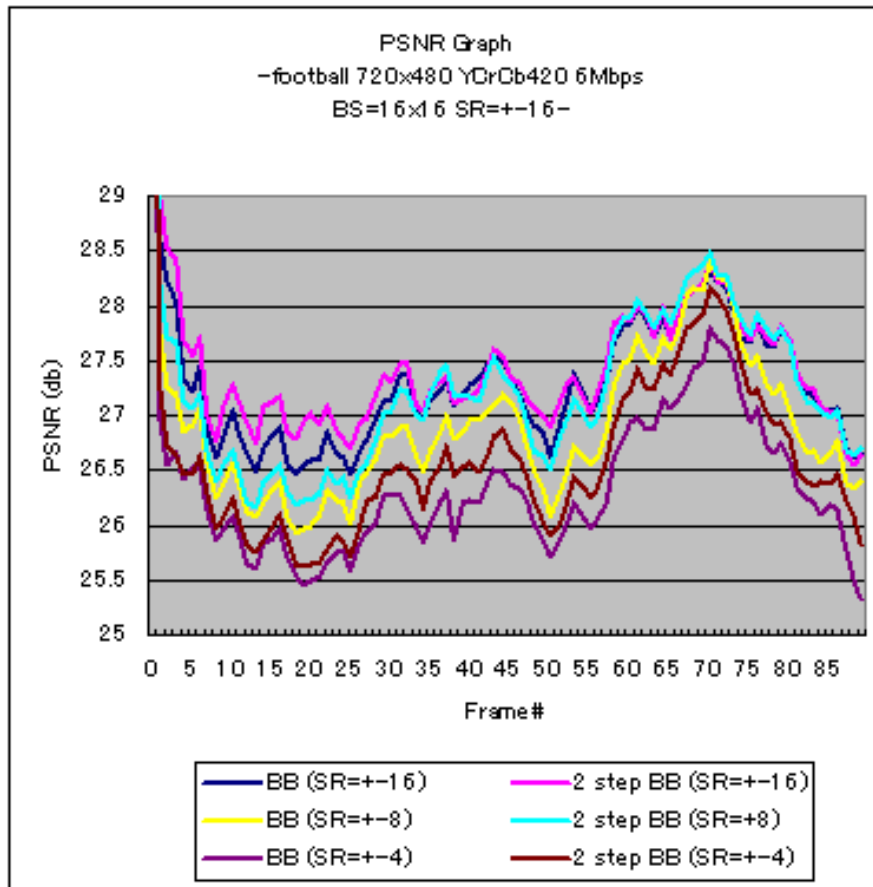


図 4.2: "football" に対する 2 段階 Band-by-band 動き推定の PSNR

画質:"football"

ほぼ全フレームで 2 段階動き推定が上回っている。まず LL バンドの動き推定は動きベクトルをプロットしてみるとほぼ良好に実際の動きを表している。ブロックサイズ 16x16 は LL2 バンドでは 4x4 となるため単なる SAD の似たブロックではなく実際の動きを表してくれているからだろう。(これがもし分解レベル 3 だったら LL3 バンドでは 2x2 となり実際の動きを表さなくなり、他バンドの推定にも悪影響を及ぼしていたに違いない。)そして他バンドの動き推定は LL バンド動きベクトルの終点をサーチの中心とするためより高確率に動き推定できる。特に LH1 バンドなどは元々はサーチレンジ ± 16 では動きを覆いきれなかったのであるが、2 段階にすることによりサーチレンジ ± 32 にあたる範囲をすべてではないが探索してくれるのでより LH1 バンドの差分画像のゲインを減らすことが出来ている。このような理由から "football" では全フレームにおいて高画質となっているのだろう。

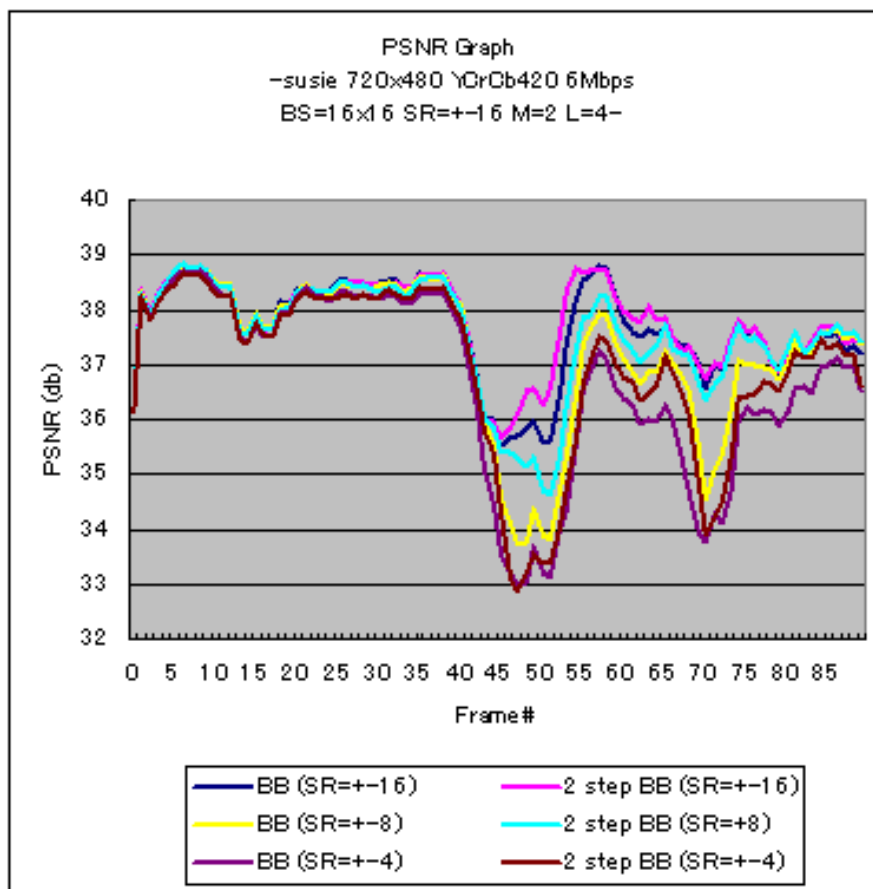


図 4.3: "susie" に対する 2 段階 Band-by-band 動き推定の PSNR

画質:"susie"

"susie" は動きの比較的小さい動画であるため提案手法はそれほど効果を発揮しない。"それほど"と述べたのは、"susie"シーケンスでは 45 番フレームから 60 番フレームにかけて susie が大きく首を振るシーンがあるのだが、この部分が大きな動きであり、提案手法が効果を発揮しているからである。このシーンでは実験した全てのサーチレンジで従来法を上回っている。

"football"・"susie"の結果から提案手法が高解像度の動きの大きい動画に強いことが示された。

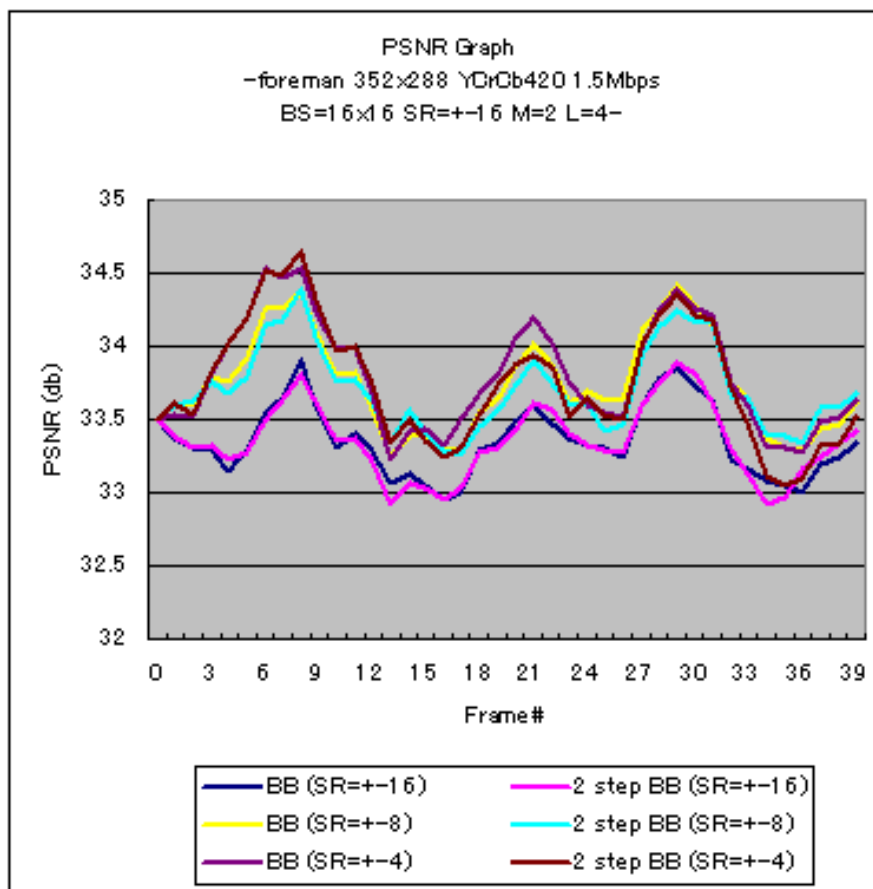


図 4.4: "foreman" に対する 2 段階 Band-by-band 動き推定の PSNR

画質:"foreman"

"foreman"シーケンスに映っているレポーターは何度も顔を振って話しているので他の試験動画に比べて動きが大きいといえば大きい。しかしながら解像度が SIF(352x288) であるため"football"・"susie"のときと同じブロックサイズ・サーチレンジにおいて動きが大きいと言えるかどうかは疑問である。

実験結果は従来法と提案手法でほぼまったく変わらなかった。サーチレンジ ± 4 のときに 0.05dB 程度提案手法が悪かったくらいであり、大きな違いではない。

"foreman"の結果から低解像度の動きの大きい動画にはほぼ影響はないことがわかった。今後は高解像度動画が動画の標準になっていくだろうが、この問題を検討する必要があるだろう。

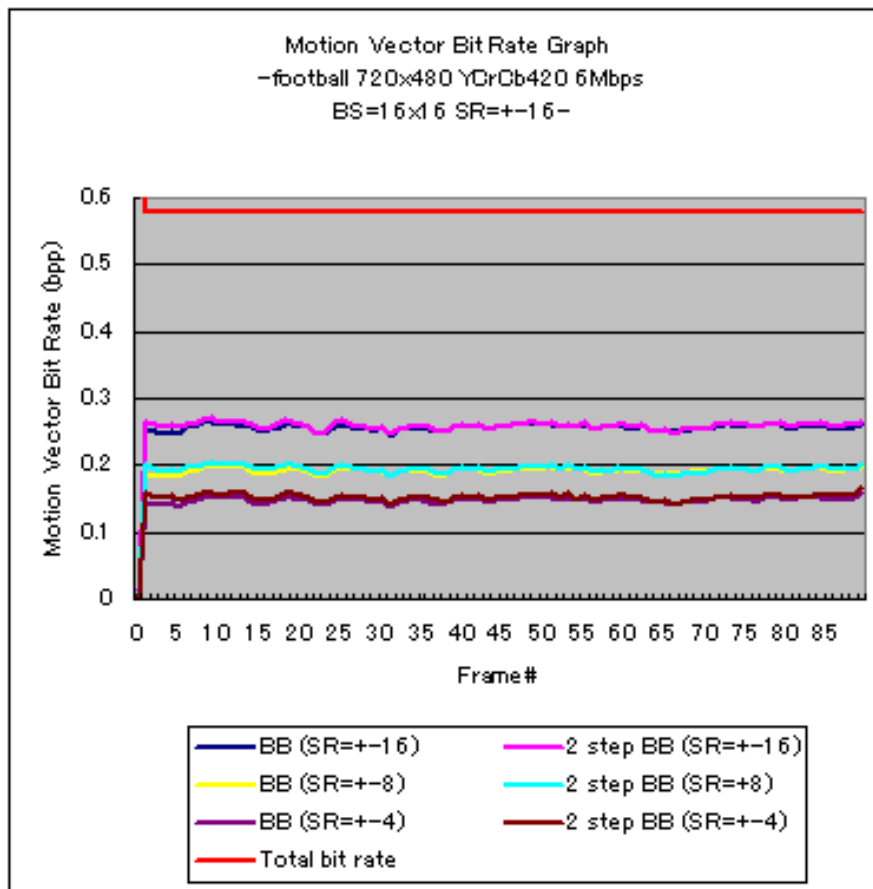


図 4.5: "football" に対する 2 段階 Band-by-band 動き推定の動きベクトルビットレート

動きベクトルビットレート

動きベクトルのビットレートはほとんど変化していない。従来大きな動きを覆うためには大きなサーチレンジが必要だったがこうすると動きベクトルの種類が増え、動きベクトルのビットレートが増えてしまう。提案した 2 段階動き推定ではサーチレンジを大きくしていないこと、また他バンドでは LL バンド動きベクトルと求めた動きベクトルの差分ベクトルを保存しているのでこのように動きベクトルの種類が増えることはない。つまり提案手法は動きベクトルのビットレートを増やすことなく大きな動きに対応しているといえる。

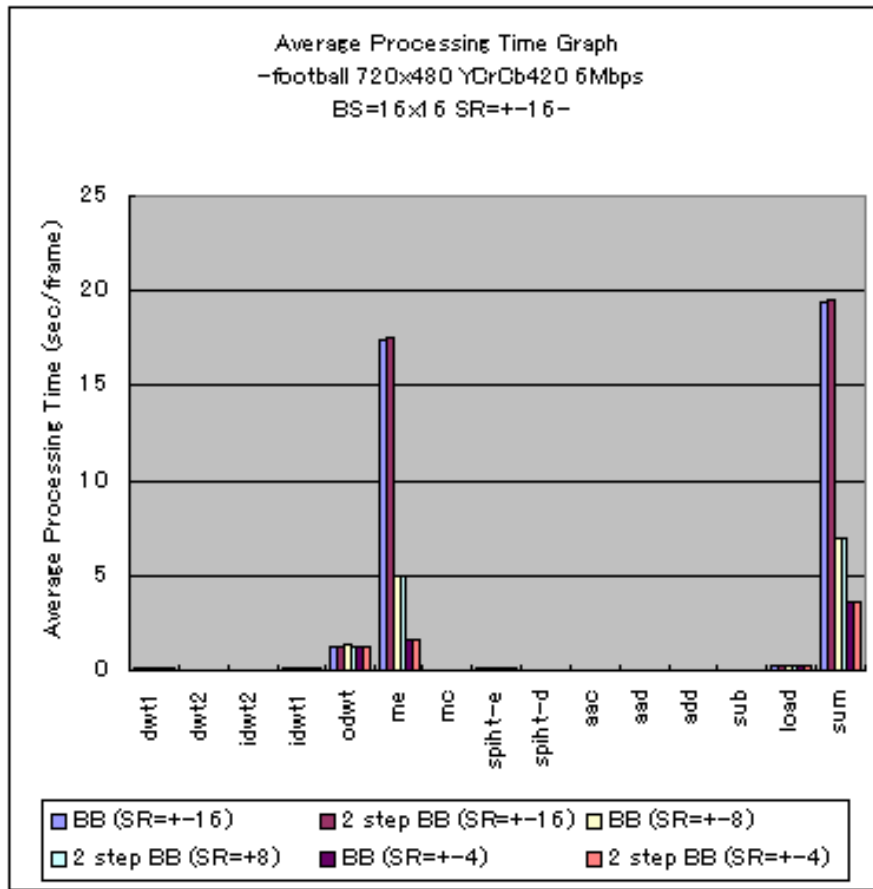


図 4.6: "football" に対する 2 段階 Band-by-band 動き推定の平均処理時間

平均処理時間

計算時間のオーバーヘッドはない。当然ながら何か特別な処理を増やしているわけではないのでオーバーヘッドが発生することはない。

4.7 結論

本章では動きベクトルの帯域間相関を用いた 2 段階 Band-by-band 動き推定を提案・評価した。これにより高解像度の動きの大きな動画に対して計算量を増やすことなく対応できるようになり画質を向上させた。

第5章 LH1バンドの係数値を用いた動的探索範囲設定

本章では前章で提案した動きベクトルの帯域間相関を利用した2段階 Band-by-band 動き推定的高速化の手法として動きベクトルの大きさと LH1 バンドの係数値の相関を用いた動的探索範囲設定を提案する。

5.1 仮説

前章までの方法ではサーチレンジは ± 16 と一定だった。そのため動きの少ない部分でも広い範囲を探索していた。そこで動きの少ない部分のサーチレンジを動的に小さくする方法を考える。

まず、以下のような仮説を立てる。

- 低周波部分はサーチレンジが小さくてもよい。

低周波部分は動き補償誤差が少ない。低周波部分は色の変化が滑らかまたは一定でありサーチレンジを小さくして近くの色の似たブロックで代用しても動き補償誤差をあまり生じないからである。

5.2 検証

前節の仮説を検証する。例として”football”16 番フレーム・”susie”46 番フレーム・”foreman”1 番フレームの DWT 係数と動きベクトルをそれぞれ図 5.1・5.2・5.3 に示す。

まず、「低周波部分」と「高周波部分」を区別するための判断基準として LH1 バンドを観察する。図 5.1・5.2 からわかるように LH1 バンドにもっとも良く高周波・低周波の差が出ている。

次に LH1 バンドと動きベクトルの関係を観察する。同様に、図 5.1・5.2 からわかるように実際に LH1 バンドにおいて低周波となっている部分の動きベクトルは大きさが小さい。

しかしながら図 5.3 に示す”foreman 352x288”シーケンスでは全く前節の仮説は成り立たない。なぜなら動きベクトルの大きさと LH1 バンドの係数値に相関性がほとんどないからである。動きベクトルの大きさが大きいのは顔全体であるのに対し、LH1 バンドにそれが高周波となって現れていない。”football 720x480”・”susie 720x480”のような高解像度自然画像では LH1 バンドに高周波がくっきり表示されるため本手法は有効であるが、低解像度動画では高周波が出にくいいためあまり役に立たない。これはマザーウェーブレットのフィルタ長なども関係してくるだろう。今後の課題とする。

次に”football 720x480”・”susie 720x480”に対し、この性質を利用して次のようなアルゴリズムを提案する。

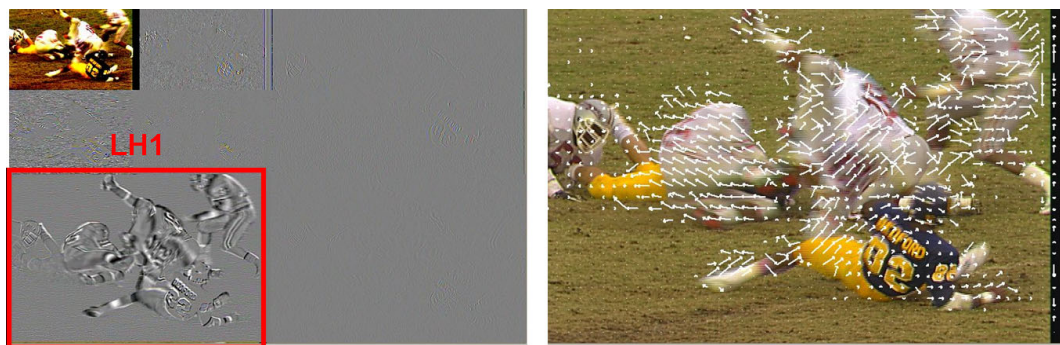


図 5.1: "football" 16 番フレームの DWT 係数 (左) と動きベクトル (右)

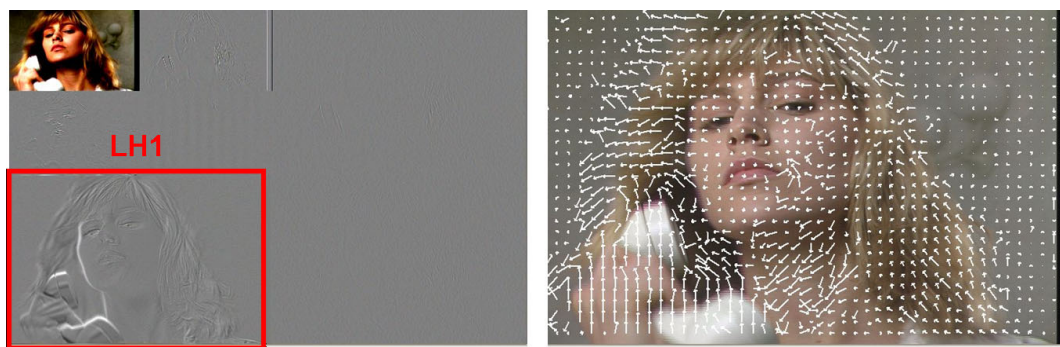


図 5.2: "susie" 46 番フレームの DWT 係数 (左) と動きベクトル (右)

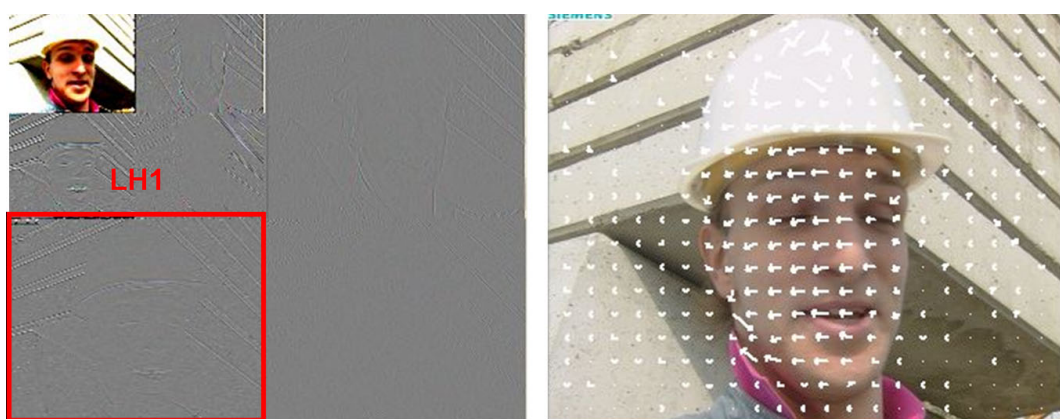


図 5.3: "foreman" 1 番フレームの DWT 係数 (左) と動きベクトル (右)

5.3 アルゴリズム

1. 初期化

サーチレンジを ± 16 に設定

```
search_range = 16;
```

2. カレントブロック (現時刻の着目ブロック) に対するカレントフレーム LH1 ブロックの絶対値平均計算

カレントブロック (左上座標 (x, y)) と空間的に同位置のカレントフレームの LH1 ブロック (左上座標 $(\frac{x}{2}, \frac{y}{2})$) に対してその係数の絶対値の平均 M_{cur} を計算。B はブロックサイズを表す。

$$M_{cur} = \frac{1}{\left(\frac{B}{2}\right)^2} \sum_{i=0}^{\frac{B}{2}-1} \sum_{j=0}^{\frac{B}{2}-1} \left| LH_{cur}^{(1)} \left(\frac{x}{2} + i, \frac{y}{2} + j \right) \right|$$

3. カレントブロックに対する参照フレーム LH1 ブロックの絶対値平均計算

カレントブロック (左上座標 (x, y)) と空間的に同位置の参照フレームの LH1 ブロック (左上座標 (x, y)) に対してその係数の絶対値の平均 M_{ref} を計算。参照フレームは ODWT 領域であるため計算方法は異なることに注意されたい。B はブロックサイズを表す。

$$M_{ref} = \frac{1}{\left(\frac{B}{2}\right)^2} \sum_{i=0}^{\frac{B}{2}-1} \sum_{j=0}^{\frac{B}{2}-1} \left| LH_{ref}^{(1)}(x + 2i, y + 2j) \right|$$

4. 動的サーチレンジ設定

カレントフレーム・参照フレームの LH1 ブロックの係数絶対値平均 M_{cur}, M_{ref} が共に閾値未満ならカレントブロックを低周波ブロックであると判断し、サーチレンジを ± 2 とする。それ以外の場合はサーチレンジを ± 16 のまま変更しない。

```
if ( M_cur < threshold && M_ref < threshold ) search_range = 2;
```

5. 変更されたサーチレンジでフルサーチを行う。

ここで参照フレームの LH1 ブロックも考慮に入れた点について説明する。カレントブロックが現時刻で低周波だとしても、1 フレーム前の参照フレームにおける同座標は高周波だった可能性がある。これは背景等が見え隠れするような画像のエッジ付近に起こりやすい。例えばカレントブロックは顔の頬の部分であるのに対し、参照フレームの同座標は背景となっていた場合である。この可能性を無視し、現時刻で低周波だからといってサーチレンジを小さくしてしまったら、参照フレームから背景ブロックを顔に動き補償しかねない。このような理由から、カレントブロックはカレントフレームと参照フレームの両方において低周波だった場合のみサーチレンジを小さくしている。

5.4 実験

CODEC は図 3.8 とする。"CIPR ITU601 football 720x480"・"CIPR ITU601 susie 720x480"・"CIPR SIF foreman 352x288"シーケンスを用い実験を行った。YCbCr4:2:0 変換を行い、ビットレートは 6Mbps (I フレームのみ 2.0bpp で符号化、それ以外は誤差画像と動きベクトルのビットレート)。またブロックサイズは 16x16、サーチレンジは ± 16 とした。

"football"の PSNR・平均処理時間をそれぞれ図 5.4・5.5、平均 PSNR を表 5.1 に、"susie"の PSNR・平均処理時間を図 5.6・5.7、平均 PSNR を表 5.2 に示す。

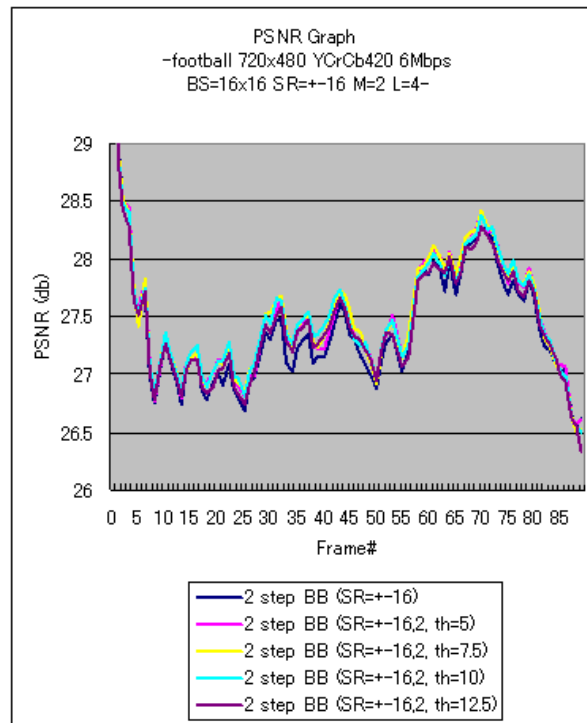


図 5.4: "football" に対する動的探索範囲 2 段階 Band-by-band 動き推定の PSNR

表 5.1: 動的探索範囲評価実験における"football"シーケンスの平均 PSNR (2-BB:2 step Band-by-band, 2-BB:2 step BB, SR:Search Range)

| Sequence Name | 2-BB | th:5.0 | th:7.5 | th:10.0 | th:12.5 |
|---------------|-------|--------|--------|---------|---------|
| football | 27.45 | 27.55 | 27.56 | 27.56 | 27.49 |

図 5.4・表 5.1 より、閾値 7.5 または 10 のときに PSNR が最大となっていることがわかる。しかも元の何も行わない 2 段階推定よりも PSNR が高い。これは動きの少ない部分をサーチレンジ ± 2 にしたことにより動きベクトルが局所化され、この部分の動きベクトルのデータ量が少なくなった分、差分画像にビットレートを割り当てられたことによる。

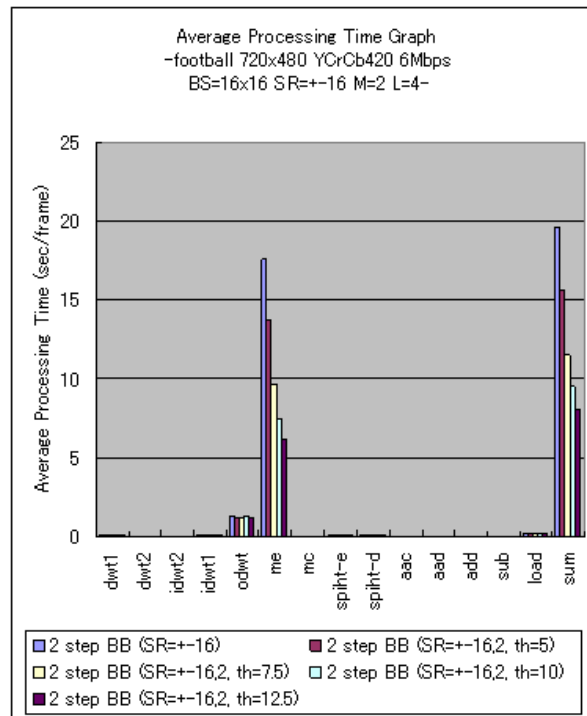


図 5.5: "football" に対する動的探索範囲 2 段階 Band-by-band 動き推定の平均処理時間

図 5.5 から閾値を上げると動き推定に要する時間が減っていくことがわかる。これは閾値を上げると低周波とみなされサーチレンジが小さくなるブロックが増えたことによる。これと図 5.4 から、"football 720x480" には閾値 10 が適していると判断し、結果的に動き推定の平均処理時間を約 42.7%に減らすことができた。

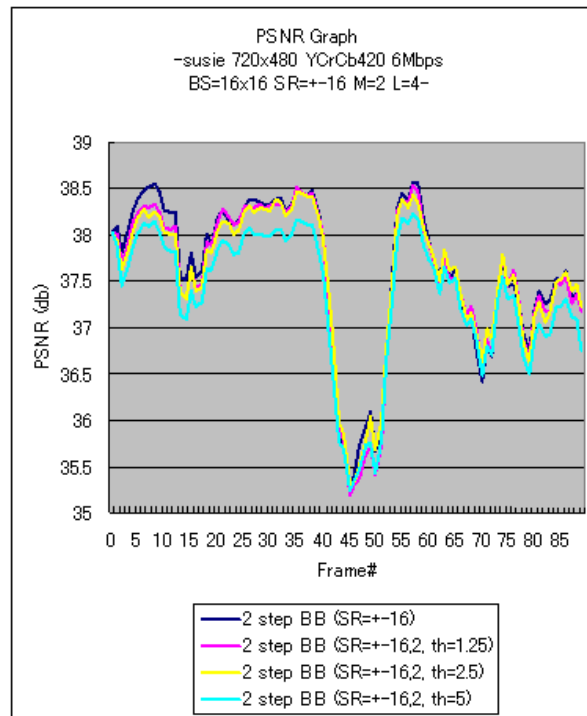


図 5.6: "susie"に対する動的探索範囲 2 段階 Band-by-band 動き推定の PSNR

表 5.2: 動的探索範囲評価実験における"susie"シーケンスの平均 PSNR (2-BB:2 step Band-by-band, 2-BB:2 step BB, SR:Search Range)

| Sequence Name | 2-BB | th:1.25 | th:2.50 | th:5.00 |
|---------------|-------|---------|---------|---------|
| susie | 37.61 | 37.56 | 37.56 | 37.35 |

図 5.6・表 5.2 より閾値を上げていくと 2.5 のときまでは一部を除いてあまり PSNR が減少していないことがわかる。そして閾値を 5 にしたとき明らかに全体的に PSNR が下がっている。

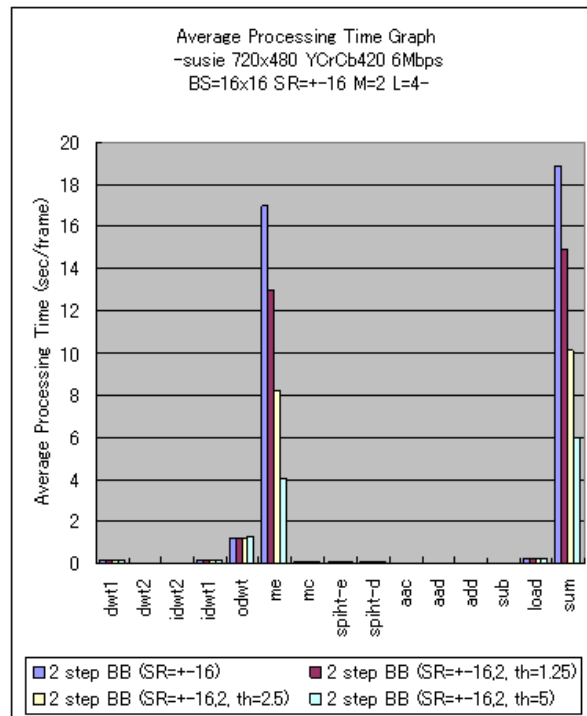


図 5.7: "susie" に対する動的探索範囲 2 段階 Band-by-band 動き推定の平均処理時間

図 5.6 から閾値 2.5 でもっとも PSNR 減少を抑え、かつ高速化できていることがわかる。そして図 5.7 とから何も行わない 2 段階動き推定の約 48.3%の計算時間に減らせていることがわかる。

5.5 結論

本章では LH1 バンドの係数値を用いた動的探索範囲設定を提案し、提案した 2 段階推定の計算量を約約 42.7%("football")、約 48.3%("susie") に減らすことができた。

第6章 結論

6.1 まとめ

本論文では Band-by-band 動き推定を高解像度動画において最適化し、動きの大きい動画を高画質かつ高速にした。

第2章 動き補償を行う最適な領域、そしてその領域における動き補償アルゴリズムの関連研究を紹介した。

第3章 静止画像は分解レベル5が最適であること、また動き補償は分解レベル2が最適であることを示しこの二つの問題を同時に解決する CODEC を提案し、0.2dB ~ 1.0dB の画質向上を得た。

第4章 動きベクトルの帯域間相関を用いた2段階 Band-by-band 動き推定を提案し、動きの大きい動画の画質をサーチレンジによるが約 0.5dB 向上させた。

第5章 LH1 バンドの係数値を用いた動的探索範囲設定を提案し、提案した2段階推定の計算量を約約 42.7%("football")、約 48.3%("susie") に減らすことができた。

従来ウェーブレット系 CODEC は動き補償に向かないといわれてきたが、その主な原因は高分解レベルにおける動き補償があまり効率的ではなかったことである。本研究ではこの問題を2段階 DWT に解決した点が最も新規性がある。

6.2 今後の課題

6.2.1 動的探索範囲設定における適応的閾値設定

今回第5章において提案した動的探索範囲設定手法は閾値を実験から求めたが、これは動画により異なった。これでは実用性がないので、動画によって適応的に閾値を求める手法が必要となる。

6.2.2 3段階動き推定

今回、2段階の動き推定により探索箇所をしぼることができた。この後はこの絞られた探索範囲の中から最適なブロックを発見することが課題となる。従来の3ステップサーチなど最急降下法に基づいた動き推定を行うことにより本手法は実用的なものになるだろう。

謝辞

相田仁教授には私が勝手に自分で決めたテーマをにもかかわらず様々なアドバイス頂きました。また研究者の心構え、論文の書き方もご教授くださいました。この2年間で学んだことを忘れずにこれからも研究の道に邁進して行きたいと思います。心から感謝致します。

本研究テーマを決めるきっかけとなった2003年度夏学期講義「動画像符号化方式」をされたワシントン大学のMing-Ting Sun教授に感謝致します。この講義でCODECの面白さを学び、このテーマに挑戦できたことを嬉しく思います。

共に厳しい2年間の研究生生活の苦しみを分かち合い、共に最後まで全力を尽くした大須賀 徹 氏、光田 智史 氏に深く感謝致します。深く高速な洞察力、大胆な行動力、特に何かの提出締め切り前の底力などは素晴らしいかと思います。共に大きく成長した楽しい2年間でした。

本研究を進めていく上で多くの方から有意義なアドバイスや御支援を頂きました。簡単ではありますがお世話になりました次の方々に感謝の意を表したいと思います。

Kultida Rojviboonchai 女史

前川 智則 氏

Ng Haung Wei 氏 清水 力 氏

富井 泰介 氏

Joseph de Bonnafos 氏

岩村 尚 氏

板谷 俊輔 氏

松井 豊和 氏

Prapatsaranon Palarp 氏 Hoang Thach 氏

岡野 諭 氏

川村 泰二郎 氏

神堀 真也 氏

柳 春楊 氏

稲垣 素子 女史

中山 早百合 女史

千葉 新吾 技術官

そして、たくさんのチャンスを与えてくれた家族に感謝します。ありがとう。

2005年1月31日

参考文献

- [1] ISO/IEC 15444-1, "JPEG2000 Coding of Still Pictures", <http://www.jpeg.org/JPEG2000.html>
- [2] ISO/IEC 15444-3, "Motion JPEG 2000", <http://www.jpeg.org/JPEG2000.html>
- [3] 小野定康、鈴木純司、"わかりやすい JPEG2000 の技術"、オーム社、2003 .
- [4] The MathWorks, "Wavelet Toolbox"、
<http://www.mathworks.de/access/helpdesk/jhelp/toolbox/wavelet/wavelet.shtml>
- [5] 芦野、森本、"ウェーブレット解析とその応用"
- [6] 芦野、"ウェーブレット解析入門"
- [7] 新島、"ウェーブレット画像解析"
- [8] 上智大学ホームページ、ウェーブレット解説（現在抹消済み）
- [9] 新井康平、"ウェーブレット解析の基礎理論"、森北出版、2000 .
- [10] 浜田望、"よくわかる信号処理"、オーム社、1995 .
- [11] Henri Poincare, "Wavelets, Filter Banks and Multiresolution Signal Processing".
- [12] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC), DRAFT ISO/IEC 14496-10 : 2002 (E).
- [13] H.264/AVC 教科書、インプレス、2004.
- [14] 藤原洋、安田浩、"ポイント図解式標準ブロードバンド + モバイル MPEG 教科書"、ASCII、2003 .
- [15] vcodex, "H.264 / MPEG-4 Part 10 White Paper", www.vcodex.com
- [16] J.M.Shapiro, "Embedded image coding using zerotrees of wavelets coefficients", IEEE Trans. Signal Processing, vol.41, pp.3445-3462, Dec.1993.
- [17] C.Valens, "Embedded Zerotree Wavelet Encoding", 1999.
- [18] A.Said and W.A.Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees", IEEE Trans. Circuits and Systems for Video Technology, vol.6, no.3, pp.243-250, June 1996.

- [19] H.W.Park, H.S.Kim, "Motion estimation using low-band-shift method for wavelet-based moving picture coding", IEEE Transaction on Image Processing, vol.9, no.4, pp.577-587, April 2000.
- [20] H.W.Park, H.S.Kim, "Wavelet-based moving-picture coding using shift-invariant motion estimation in wavelet domain", Signal Processing: Image Communication, vol.16, no.7, pp.669-679, April 2001.
- [21] Hamed Sari-Sarraf and Dragana Brzakovic, "A shift-invariant discrete wavelet transform".
- [22] Suxia Cui, "Motion estimation and compensation in the redundant wavelet domain", A dissertation submitted to the faculty of Mississippi state university, August 2003.
- [23] T. Naveen and J. W. Woods, "Motion Compensated Multiresolution Transmission of High Definition Video", IEEE Trans. Circuits and Systems for Video Technology, Vol. 4, No. 1, pp. 29-41, February 1994.
- [24] T. Naveen and J. W. Wood, "Rate Constrained Multiresolution Transmission of Video", IEEE Trans. Circuits and Systems for Video Technology, Vol. 5, No.3, pp. 193-206, June 1995.
- [25] Gregory J.Conklin, Sheila S.Hemami, "Multi-resolution Motion Estimation", Visual Communication Lab, School of Electrical Engineering, Cornell University, Ithaca, NY 14853.
- [26] Arvind Sudarsanam, Mayur Srinivasan, "Multiresolution Motion Estimation", A project Report Presented in Partial Fulfillment of the Requirements for EEE 508 - Digital Image Processing and Compression, Arizona State University, Spring 2002.
- [27] Yiannis Andreopoulos, Adrian Munteanu, Geert Van der Auwera, Peter Schelkens and Jan Cornelis, "Wavelet-based fully-scalable video coding with in-band prediction", Proc.3rd IEEE Benelux Signal Processing Symposium (SPS-2002), Leuven, Belgium, March 21-22, 2002.
- [28] Claudia Mayer, Simon Albert, "In-band motion estimation and compensation for spatially scalable predictive wavelet video coding", Institute of Communications Engineering Aachen University (RWTH).
- [29] Jeongnam Youn, Ming-Ting Sun, Fellow, IEEE, and Chia-Wen Lin, "Motion Vector Refinement for High-Performance Transcoding", 1999 IEEE.
- [30] Yonghui Wang, Suxia Cui and James E.Fowler, "3D Video coding using redundant-wavelet multi-hypothesis and motion-compensated temporal filtering", In Proceedings of the IEEE International Conference on Image Processing, Barcelona, Spain, September 2003, vol.2, pp.755-758.
- [31] Suxia Cui, Yonghui Wang, and James E.Fowler, "Mesh-based motion estimation and compensation in the wavelet domain using a redundant transform", In Proceedings of the IEEE International Conference on Image Processing, Rochester, NY, September 2002, vol.1, pp.693-696.
- [32] Jens-Rainer Ohm, Mihaela van der Schaar, "Wavelet Video Coding - Principles and
- [33] Anil Aksay, "Motion Wavelet Video Compression", A thesis submitted to the graduate

- [34] G.Van der Auwera, A.Munteanu, G.Lafruit and J.Cornelis, "Video coding based on motion
- [35] Thomas Wiegand, Bo Lincoln, and Bernd Girod, "Fast search for long-term memory motion-compensated prediction", in the proceedings of the IEEE International Conference on Image Processing October 4-7, 1998, Chicago, USA.
- [36] James E.Fowler, "QccPack : Quantization, compression, and coding utilities", 1997-2004.
- [37] Amir Said, "FastAC: Arithmetic Coding Implementation", <http://www.cipr.rpi.edu/said/FastAC.html>
- [38] 上村、張、吉川、"全域通過フィルタを用いた画像ウェーブレット符号化"、電子情報通信学会論文誌、2001/1 Vol. J84-A No.1.
- [39] 大巻、董、藤田、尾上、白川、"離散ウェーブレット変換によるリアルタイムビデオ符号化器"、大阪大学工学研究科情報システム工学専攻.
- [40] Ming-Ting Sun, "Digital Video Coding", 2003 年度夏学期授業資料.
- [41] 小早川 知昭、相田 仁、斉藤 忠夫、"MPEG 符号化方式の拡張による 3 次元動画像通信"、東京大学大学院修士論文、1997.2.
- [42] 香山 美奈子、相田 仁、斉藤 忠夫、"領域マッチングによる視差推定を用いた多眼画像符号化方式"、東京大学大学院修士論文、2000.2.
- [43] 前川 智則、相田 仁、"インターネットにおけるエンド・エンドでの複数パス利用プラットフォーム"、東京大学大学院修士論文、2004.1.

発表文献

[1] 佐藤 大将、相田 仁、"ウェーブレットの位置情報保存性を利用した高精度動き推定"、電子情報通信学会 画像工学・映像情報メディア学会 メディア工学合同研究会、2004.9.