

平成19年度 修士論文

並列木複素ウェーブレット変換を用いた
静止画像符号化の研究

Research on Still Image Coding
Using Dual-Tree Complex Wavelet Transform

指導教授 渡辺 裕 教授

2008年3月

早稲田大学大学院 国際情報通信研究科
国際情報通信学専攻

4206A143-9

高 橋 良 知

Yoshitomo TAKAHASHI

目 次

第 1 章 序論	1
1.1 研究の背景	1
1.2 本研究の目的	2
1.3 本論文の構成	4
第 2 章 静止画像の変換符号化の現状	6
2.1 まえがき	6
2.2 国際標準静止画像符号化	6
2.2.1 変換符号化の基本構成	6
2.2.2 JPEG	7
2.2.3 JPEG 2000	10
2.3 離散ウェーブレット変換	14
2.3.1 連続ウェーブレット変換	14
2.3.2 パラメータの離散化	15
2.3.3 多重解像度解析の原理	16
2.3.4 フィルタバンクによる実装	18
2.3.5 完全再構成	20
2.3.6 2次元離散ウェーブレット変換	22
2.4 離散ウェーブレット変換の問題点	23
2.4.1 シフト不変性の欠如	23
2.4.2 低性能な方向分離特性	25
2.5 離散ウェーブレット係数の SPIHT 符号化と符号化性能	26
2.6 むすび	31
第 3 章 並列木複素ウェーブレット変換による新しい静止画像符号化	33
3.1 まえがき	33
3.2 複素ウェーブレット変換の種類	33
3.3 ヒルベルト変換と解析信号	35
3.4 複素ウェーブレットとフーリエ変換	36
3.5 並列木複素ウェーブレット変換	37
3.5.1 並列木フィルタバンクによる実装	37
3.5.2 ヒルベルト変換対のフィルタ設計	38

3.5.3	並列木フィルタバンクの制約	41
3.5.4	2次元並列木複素ウェーブレット変換	43
3.6	並列木複素ウェーブレット変換の特性	43
3.6.1	近似的シフト不変性の成立	43
3.6.2	高性能な方向分離特性	46
3.7	並列木複素ウェーブレット係数の冗長性	47
3.7.1	複素係数の絶対値の冗長性	48
3.7.2	複素係数の位相の冗長性	49
3.7.3	並列木複素ウェーブレット係数の SPIHT 符号化と符号化性能	52
3.8	冗長変換の疎表現	53
3.8.1	冗長変換と疎表現の概念	53
3.8.2	疎表現の理論	55
3.8.3	マッチング追跡近似解法	56
3.8.4	基底追跡近似解法	57
3.9	むすび	59
第4章	並列木複素ウェーブレット変換の疎表現における係数選択手法と係数予測手法	60
4.1	まえがき	60
4.2	並列木複素ウェーブレット変換の疎表現を用いた静止画像符号化システム	60
4.3	疎表現近似解法の考察と係数選択手法の提案	61
4.3.1	Noise Shaping 法	61
4.3.2	Noise Shaping 法の問題点	63
4.3.3	Noise Shaping 法の疎表現的考察	63
4.3.4	Noise Shaping 法における新しい係数選択手法の提案	65
4.4	疎表現係数列を用いた静止画像符号化実験	67
4.4.1	Noise Shaping 法のパラメータと疎表現の性能	67
4.4.2	非ゼロ係数へのエネルギー集約度による提案手法の性能	68
4.4.3	全符号量による提案手法の性能	68
4.4.4	疎表現を用いた静止画像符号化の考察	71
4.5	疎表現係数列における係数予測手法	75
4.5.1	疎表現係数の絶対値予測手法	75
4.5.2	予測手法の評価実験	77
4.6	むすび	78
第5章	結論	79
5.1	総括	79
5.2	今後の課題	80
5.2.1	非冗長複素ウェーブレット変換の検討	80
5.2.2	疎表現の近似解法の検討	80

5.2.3	疎表現係数列のベクトル量子化とコンテキストモデリングの検討 . . .	80
5.2.4	疎表現係数列における位相予測手法の検討	81
謝辞		82
参考文献		83
付 録 A 値域と零空間		89
図一覧		90
表一覧		93
研究業績		94

第1章

序論

1.1 研究の背景

近年、急速なデジタル技術の発達と普及により、音楽・画像・映像などのマルチメディアコンテンツが、我々にとって極めて身近な存在となっている。FTTH (Fiber To The Home) や ADSL (Asymmetric Digital Subscriber Line) などのブロードバンドインターネットサービスの世帯普及率は 50 [%] を超え、大容量のデータ通信の実現により GyaO⁽¹⁾ のようなユーザの要求する映像コンテンツを即座に配信する VOD (Video On Demand) サービスが一般的となってきた。また、画像共有サービス Flickr⁽²⁾ や映像共有サービス YouTube⁽³⁾、ニコニコ動画 (RC2)⁽⁴⁾ のように画像・映像コンテンツを不特定多数で共有するサービスも注目を集めている。さらに、HDD (Hard Disk Drive) の大容量化に伴いホームサーバーに映像コンテンツを蓄積し、フレキシブルな映像視聴が可能となるサーバー型放送のサービスがまもなく開始されようとしている。

このような多様なサービスの中で、映像コンテンツ自体は高解像度化と携帯端末対応へと発展している。撮像装置では、1000 万画素を超える解像度で撮影可能なデジタルスチルカメラや 60 [fps] で映像を記録できるデジタルビデオカメラがコンシューマー向けに登場している。表示装置ではテレビの地上波アナログ放送が終了する 2011 年 が近づくにつれ、フル HD (High Definition) 規格対応の液晶テレビやプラズマテレビが普及してきている。また、世界初となる 100 万 : 1 以上の高コントラスト比を持ち、厚さ 3[mm] の 11 型有機 EL (Electro-Lumminescence) テレビも大画面化を目指し開発が続けられている⁽⁵⁾。NHK では 8000 × 4000 画素サイズのスーパーハイビジョンを 2020 年までに一般家庭に普及させることを目的として研究をしており⁽⁶⁾、今後も映像コンテンツの高解像度化は持続すると考えられる。携帯端末では、ワンセグ放送の開始に伴い携帯端末で映像を視聴する機会が増加している。また、PSP⁽⁷⁾ や ニンテンドー DS⁽⁸⁾ のような携帯ゲーム端末へ動画像を保存し移動中に視聴することも可能である。

一方、デジタルコンテンツの効率的な伝送・蓄積を行うために必要不可欠な符号化技術は、静止画像符号化技術として 2000 年に JPEG 2000 が、動画像符号化技術として 2003 年に MPEG-4 Part10 AVC (Advanced Video Coding) が、ISO/IEC JTC 1/SC 29 において国際標準化され、現在も日々進歩を遂げている。MPEG-4 Part10 AVC は、ITU-T で

は H.264 として標準化されており, MPEG-2 と比較して約 2 倍の高い圧縮率を実現した符号化方式である. 一方で, JPEG 2000 は, 可逆符号化 (lossless coding) および非可逆符号化 (lossy coding) の双方を実現する符号化方式である. JPEG 2000 は, 現在デジタルスチルカメラやインターネットで普及している JPEG (Joint Photographic Experts Group) よりも高い圧縮効率を実現すると共に, 品質スケーラビリティと空間スケーラビリティが実現可能である. 符号化データのスケラビリティとは, 符号化データの一部のみを復号することで, 任意の品質のデータ再生や, 低解像度の画像を取得することを可能とする仕組みである.

現在までに, 静止画像符号化方式として JPEG と JPEG 2000 が国際標準化されてきたが, これらの方式は“直交変換を用いた変換符号化”という点で共通している. 直交変換として, JPEG では離散コサイン変換 (DCT : Discrete Cosine Transform), JPEG 2000 では離散ウェーブレット変換 (DWT : Discrete Wavelet Transform) が採用されている. 直交変換された係数は冗長性が削除され, 特定の係数にエネルギーを集中できる. また, 直交変換は完備な基底 (complete basis) を用いるため変換前後のデータ数が不変である. このため, 直交変換は非冗長変換 (non-redundant transform) である.

一方で, 基底数が原信号のサンプル数よりも多い過完備な基底 (overcomplete basis) を用いた冗長変換 (redundant transform) と呼ばれる変換がある. このため, 冗長変換は直交変換になり得ないが, 変換後のデータに冗長性を持たせることで非冗長変換では実現できない特性をもつことができる. たとえば, ダウンサンプリング処理を行わない DWT である離散定常ウェーブレット変換 (SWT : Stationary Wavelet Transform) は変換後の冗長性より, DWT で失われるシフト不変性を成立させることができる.

また, 画像処理分野では“方向分離特性をもつ変換”が注目されている. このような変換は一般的に冗長変換であり, 最も有名な変換として Curvelet 変換がある. 本研究で用いる並列木複素ウェーブレット変換 (DTCWT : Dual-Tree Complex Wavelet Transform) も同様の特性をもつ変換である. 方向分離特性をもつ変換は, 画像信号中に含まれるエッジ等の曲線を 2 次元で定義される方向基底を用いて表現する変換である. 方向基底を用いて 2 次元構造を高い精度で近似するため, DWT のような方向分離特性の乏しい変換と比較し, 雑音除去や特徴抽出などのアプリケーションにおいて高い性能を示している. しかし, 方向分離特性をもつ変換は変換後のデータ数が増加するため, この変換を用いた静止画像符号化の研究はほとんどされていないのが現状である.

このような研究背景において, 本研究では, 国際標準符号化において用いられている直交変換ではなく, 方向分離特性と冗長変換という特性をもつ DTCWT を用いた新しい静止画像符号化の実現を目的とする.

1.2 本研究の目的

本研究は, DTCWT を用いた静止画像符号化の実現を目的とする. 国際標準符号化方式では直交変換に基づく符号化方式が採用されてきた. 直交変換は次のような特性がある.

表 1-1: 並列木複素ウェーブレット変換と離散ウェーブレット変換の違い

	並列木複素ウェーブレット変換	離散ウェーブレット変換
変換	周波数変換	周波数変換
変換手法	冗長変換	非冗長変換 (直交または双直交)
基底	過完備基底	完備基底
データ数 (2 次元)	変換後に 4 倍に増加	変換前後で不変
変換係数	複素係数	実数
シフト不変性	成立	不成立
方向分離特性	$\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$	$0^\circ, \pm 45^\circ, 90^\circ$

- 直交基底を用いるため、変換係数間の冗長性が除去されている。
- 基底間に依存関係がないため、変換係数に対する量子化処理の影響が理解しやすい。
- 完備な基底を用いた変換であるため、変換の前後においてデータ数が不変である。

直交変換のこのような特性は、画像符号化において好都合であるため、これまで直交変換に基づく画像符号化の研究が多くなされてきた。これに対して、本研究は直交変換ではない DTCWT に基づく画像符号化を実現する。直交変換の一種である DWT と DTCWT の違いを表 1-1 に示す。DTCWT と DWT はともに周波数変換であるが、その変換特性は大きく異なっている。まず、DTCWT は冗長変換であり、過剰な基底を用いて信号を表現する。冗長変換は直交変換と比較して次のような特性となる。

- 過完備な基底を用いるため、変換係数間に冗長性が残っている。
- 基底間に依存関係があるため、変換係数に対する量子化処理の影響が不明確である。
- 過完備な基底を用いた変換であるため、変換後のデータ数が増加する。

一言でいえば、DTCWT はフーリエ変換 (FT: Fourier Transform) の正弦波をウェーブレットに置き換えた変換である。そのため、DTCWT は FT に類似した特性をもっている。まず、DTCWT 係数は位相情報をもつ複素数であり、シフト不変性が成立する。これに対して、DWT は実係数であり、シフト不変性は不成立である。次に、2 次元 DTCWT は $\pm 15^\circ, \pm 45^\circ, \pm 75^\circ$ の 6 方向のエッジを異なるサブバンドへ分離する変換である。これに対して、2 次元 DWT は $0^\circ, \pm 45^\circ, 90^\circ$ の 3 方向のエッジしか分離できず、 $\pm 45^\circ$ 方向のエッジを同一サブバンドへ混在してしまう方向分離特性の乏しい変換である。

以上のような DTCWT の特性が画像符号化に与える影響を述べる。まず、冗長変換である DTCWT は変換後のデータ数が増加する。符号化するデータ数が増加すれば符号量も増加する。この点において、冗長変換は直交変換よりも不利であり、あまり研究がなされていない原因でもある。しかし、DTCWT は次の 2 点において直交変換よりも画像符号化に適

している。まず、冗長変換の多様な信号表現である。過完備な基底を用いる冗長変換では信号の表現が無数に存在する。この無数に存在する表現の中に疎表現 (sparse representation) と呼ばれる表現がある。疎表現とは、信号を最少数の基底で表現したものである。言い換えれば、信号を最少数の非ゼロ係数で表現した係数列であり、この係数列は極めて少数の非ゼロ係数と多数のゼロ係数からなる。疎表現係数列は少数の非ゼロ係数に最も多くのエネルギーを集約した係数列であり、DWT のような直交変換と比較して、1 係数あたりのエネルギー集約率が非常に高い。

次に、DTCWT の方向分離特性である。画像信号は平坦領域とエッジ領域から構成されていることが一般的である。このとき、エッジ領域はあらゆる方向のエッジを含んでいる可能性がある。そこで、エッジを効率的に表現するためには、その方向に沿った方向基底が必要となる。しかし、DWT のような直交変換はこのような方向基底から構成されていない。これに対して、DTCWT は 30° 間隔の 6 方向の方向基底から構成されている。このため DWT などと比較し、エッジ領域の効率的な表現が可能である。

本研究は、以上のような方向分離特性と冗長変換という特性をもつ DTCWT の疎表現に基づく画像符号化を実現することを目的とする。そして、DWT の SPIHT (Set Partitioning In Hierarchical Tree) 符号化と比較し、高い符号化効率の実現を目指す。

1.3 本論文の構成

以下に本章以降の構成を示す。

第1章 本章であり、研究の背景およびその目的について述べている。

第2章 静止画像の変換符号化の現状について述べる。まず、変換符号化を用いた国際標準画像符号化方式である JPEG と JPEG 2000 の概要を述べる。次に JPEG 2000 の変換方式である離散ウェーブレット変換について、フィルタバンクによる実装方法、完全再構成条件や 2 次元への拡張方法について述べる。そして、離散ウェーブレット変換の問題点を示す。最後に、離散ウェーブレット変換の画像符号化方式である SPIHT 符号化について述べ、その符号化性能を示す。

第3章 並列木複素ウェーブレット変換による新しい静止画像符号化について述べる。まず、2 種類の複素ウェーブレット変換を紹介し、ヒルベルト変換について述べる。そして、複素ウェーブレットとフーリエ変換の類似性を述べる。次に、並列木複素ウェーブレット変換のフィルバンクの構造、フィルタの設計方法や 2 次元への拡張方法について述べる。そして、並列木複素ウェーブレット変換の特性と係数間の冗長性を示し、この係数列からは高い符号化効率が得られないことを確認し、疎表現の必要性を述べる。最後に、疎表現の概念や理論を述べ、疎表現を求める近似手法について述べる。

第4章 並列複素ウェーブレット変換の疎表現を用いた静止画像符号化の検討を行う。まず、本研究において想定している並列木複素ウェーブレット変換の疎表現係数列を用いた符号化システムの概要を述べる。次に、雑音変形手法である Noise Shaping 法を説明し、その問題点を述べる。そして、Noise Shaping 法を疎表現近似解法の観点から考察し、新しい係数選択手法を提案する。実験により、提案手法の有効性と疎表現係数列を用いた静止画像符号化の優位性を示す。最後に、疎表現係数列における係数予測手法を提案し、実験によりその性能を示す。

第5章 本研究の総括と今後の課題について述べる。

第2章

静止画像の変換符号化の現状

2.1 まえがき

本章では静止画像の変換符号化の現状について述べる。まず、変換符号化である国際標準画像符号化方式 JPEG と JPEG 2000 の特徴と技術を述べる。次に、JPEG 2000 の変換手法である DWT の定義や実装方法を述べ、1次元 DWT を2次元へ拡張する方法を述べる。そして、DWT の問題点であるシフト不変性の欠如と低性能な方向分離特性を示す。最後に、JPEG 2000 よりも高い符号化効率を与える SPIHT 符号化について述べ、その符号化性能を示す。本研究では、DWT の画像符号化として SPIHT 符号化を用いる。

2.2 国際標準静止画像符号化

本節では、国際標準の静止画像符号化方式である JPEG と JPEG 2000 について述べる。これらの符号化方式は要素技術やターゲットアプリケーションに違いはあるが、直交変換を用いた変換符号化であるという点で共通である。

2.2.1 変換符号化の基本構成

画像圧縮は画像情報をより少ないデータに変換する手法である。一般的な画像符号化では非可逆符号化が用いられる。非可逆符号化とは復元データが完全に元に戻らない方式であり、情報源に対する歪みを許容することにより、大きな圧縮率が得られる利点がある。これに対して、可逆符号化は圧縮されたデータを伸張した際に完全に元に戻る方式である。可逆符号化として、ハフマン符号化、Lempel-Ziv 符号化、算術符号化などが良く知られている。この方式はシンボルの生起確率の偏りを利用し、偏りに応じた符号を与えることで全体としてデータ長を短くすることができる。可逆符号化はエントロピー符号化とも呼ばれる。

一般的に、非可逆符号化は変換符号化である。変換符号化とは符号化プロセスに画像空間から周波数空間への変換処理を含む方式である。自然画像のような画像信号は隣接画素の相関が高いという性質がある。そこで、画像信号を周波数成分に変換すると低域周波数にデータが偏り、高域周波数に対応する変換係数の分散は小さくなる。これより、高域係数に割り当てるビット数が削減され、全体として変換する前よりも少ないビット数で画像を符号化で

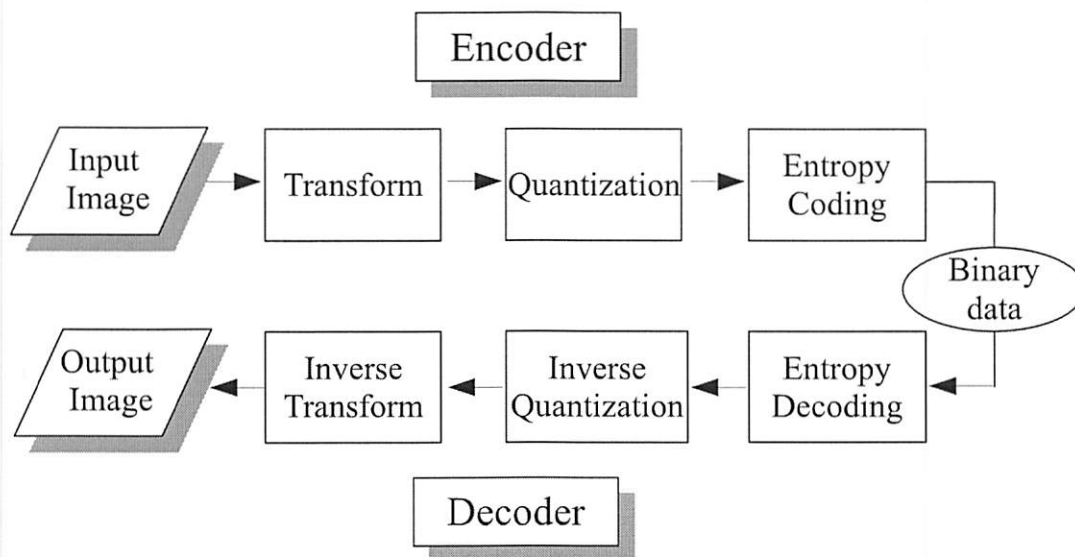


図 2-1: 変換符号化の符号化器と復号化器の構成

きる。その後、変換係数は量子化、エントロピー符号化されて伝送される。量子化とは、入力データを区切り、代表値を決定する処理である。量子化を行うことで入力データと代表値の間に誤差が生じ、復号画像に劣化が生じる。このため非可逆となる⁽⁹⁾。変換符号化の処理フローを図 2-1 に示す。

静止画像符号化方式は現在までに JPEG と JPEG 2000 の 2 つが標準化されている。JPEG と JPEG 2000 の基本構成はともに図 2-1 であり、周波数変換として直交変換を用いている。

2.2.2 JPEG

JPEG の特徴

JPEG はカラー静止画像を対象とした国際標準の圧縮符号化技術である。Joint Photographic Experts Group の略であり、ISO/IEC/JTC1/SC/WG1 と ITU-T SG 8 CCIC (Common Component for Image Communication) の共同グループとして 1986 年に設立された。この標準化作業は 1994 年に終了している。JPEG はフルカラーあるいはモノクロームのグレイスケールの自然画像を扱う種々のアプリケーションで幅広く利用できるように以下の特徴を持っている。

2つのモード JPEG にはシーケンシャルモードとプログレッシブモードがある。シーケンシャルモードは画像をラスタ順に復号するモードである。プログレッシブモードは大まかな画像を再生した後に、徐々に精細度を向上させるモードである。これらのモードはアプリケーションにより使い分けることができる。

画質の選択性 圧縮に用いるパラメータを調節して、画像品質とファイルサイズのトレードオフの中からユーザの要望に応じたものを選び出すことができる。

可逆と非可逆のサポート 非可逆符号化により、情報損失の伴った圧縮率の高い符号化に加え、DPCM (Differential Pulse Code Modulation) を用いた可逆符号化方式により、情報の劣化がない完全な符号化も可能である。

高能率符号化方式 カラー画像伝送のための通信コストを軽減できる。また、カラー画像蓄積のためのメモリ量を軽減できる高能率な符号化である。

これらの特徴より、JPEG はカラー静止画像を扱う代表的なアプリケーションであるデジタルスチルカメラのほとんどで標準仕様となっており、自然画像の圧縮に最もよく使われる。JPEG の性能は、一般的に 10 : 1 ~ 20 : 1 を視覚的な劣化なしに達成できる。また、多少の劣化を許容すれば 30 : 1 ~ 50 : 1 の圧縮率が得られる。JPEG は画像を 8×8 のブロックに分割し、ブロックごとに処理を行う。そのため低ビットレートでは、ブロック境界にブロック・ノイズと呼ばれる符号化雑音が発生する。また、エッジなどの急峻な輝度変化付近では波状のモスキート・ノイズと呼ばれる符号化雑音が発生する。これらの符号化雑音は JPEG 特有の雑音であり主観品質を下げる原因となる。

JPEG の技術

JPEG で用いられている技術を図 2-1 に対応させ簡単に述べる。まず、JPEG は直交変換である DCT を用いた周波数変換を行う。最も高い圧縮効率を与える直交変換として、カルーネンレーベ変換 (KLT : Karhunen-Loève Transform) が知られている。KLT は入力信号との自己相関に基づき変換基底を決定する。このため、他の直交変換に比べ、少数の変換係数にエネルギーを集中させることができる。しかし、KLT は信号ごとに基底を求める必要があり、基底情報を受信側に送らなければならないため、画像符号化において実用的な直交変換ではない。そのため、JPEG では KLT とほぼ同等の性能を示し、LSI (Large Scale Integration) 化が容易な高速算法が存在する DCT が採用されている。DCT は N サンプルの信号を対称的に折り返した $2N$ サンプルに対して離散フーリエ変換 (DFT : Discrete Fourier Transform) したものと等価である。入力信号 $f(x)$ に対する 1 次元 DCT (1D DCT) の順変換と逆変換は式 (2.1)、式 (2.2) となる。

$$F(u) = \sqrt{\frac{2}{N}} C(u) \sum_{x=0}^{N-1} f(x) \cos \frac{\pi(2x+1)u}{2N} \quad (2.1)$$

$$f(x) = \sqrt{\frac{2}{N}} \sum_{u=0}^{N-1} C(u) F(u) \cos \frac{\pi(2x+1)u}{2N} \quad (2.2)$$

ここで、

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}} & (u = 0) \\ 1 & (u \neq 0) \end{cases}$$

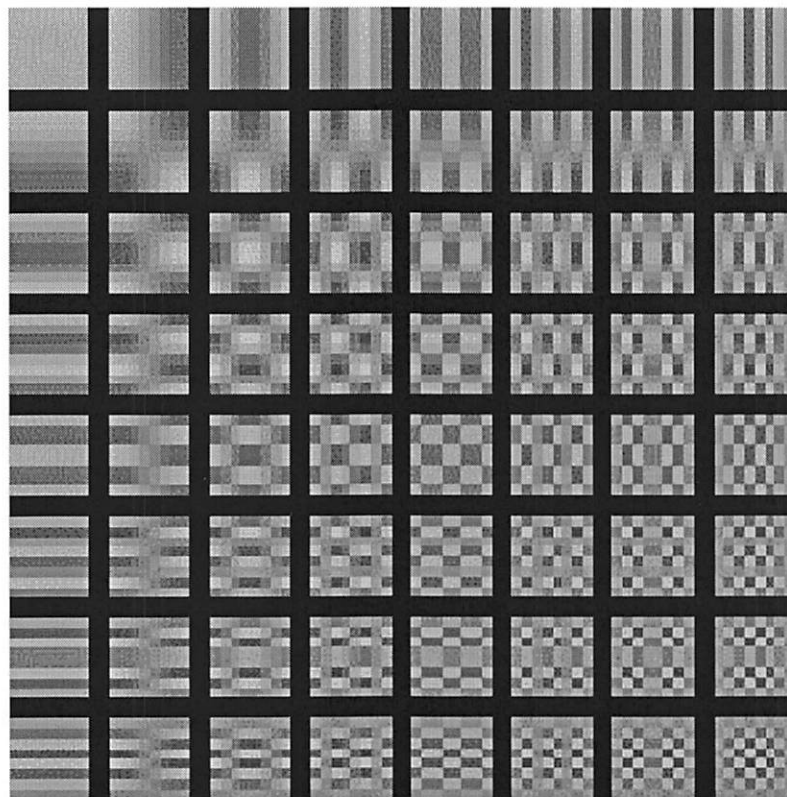
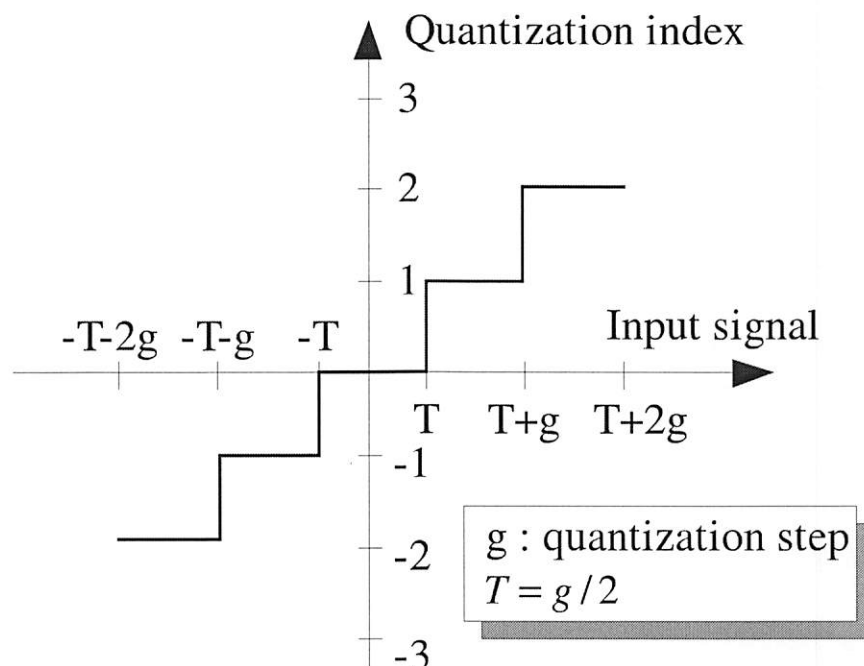


図 2-2: 2次元離散コサイン変換の基底ベクトル (文献⁽⁹⁾引用)

2次元 DCT (2D DCT) は 1D DCT を横方向、縦方向の順に適用することで実現される。JPEG では 8×8 ブロックごとに 2D DCT が適用され、その基底ベクトルは図 2-2 のようになる。

続いて、 8×8 の変換係数はジグザグ・スキャンされスカラー量子化される。このときの量子化器は図 2-3 のような $T = g/2$ のミッドトレッド型である。 8×8 ブロック内の各係数は量子化テーブルにより与えられる量子化ステップを用いて量子化される。量子化テーブルの例を表 2-1 に示す。

量子化後の変換係数はエントロピー符号化される。エントロピー符号化は DC (Direct Current) 係数と AC (Alternating Current) 係数に分けて、ハフマン符号化が適用される。ただし、拡張機能として算術符号化も可能である。まず、DC 係数の符号化では、直前ブロックの DC 成分との差分が符号化される。DC 差分はその値によってグループ番号と固定長符号により表現され、グループ番号はハフマン符号化される。AC 成分の符号化では、ジグザグ・スキャン順にゼロ係数の RUN と非ゼロ係数のレベルが符号化される。まず、RUN をカウントし、15 個以上のゼロ RUN に対しては ZRL (Zero Run Length) 符号を繰り返し用いる。非ゼロ係数のレベルは DC 差分と同様にグループ番号と固定長符号により表現する。ゼロ RUN とグループ番号は 2次元ハフマン符号を用いて符号化される。2次元ハフマン符号は 2つの事象の組み合わせに対して 1種類のハフマン符号を割り当てる手法である。

図 2-3: ミッドトレッド型量子化器 (文献⁽⁹⁾ 参照)

2.2.3 JPEG 2000

JPEG 2000 の特徴

JPEG 2000 は ISO/IEC JTC1/SC29/WG1 により 1997 年 11 月に標準化が開始され、2000 年 12 月に標準化された静止画像符号化の国際標準である。JPEG 2000 の標準化が開始された 1997 年では、現在のような高速インターネット回線が存在していなかった。そのため、狭帯域で画像を伝送するために圧縮率を上げて JPEG ファイルのサイズを小さくすることが頻繁に行われていた。しかし、JPEG では圧縮率を上げていくとブロック・ノイズやモスキート・ノイズが大きくなり、劣化が目立つようになる。そこで、JPEG 2000 は JPEG よりも高圧縮率を実現することで、このような劣化を防ぐことが目的にあった。また、圧縮率の向上以外にも JPEG 2000 は JPEG にはない多くの機能を持っている⁽¹⁰⁾。

高い圧縮効率 すべての圧縮率において JPEG を 20～30 [%] 上回る。特に高圧縮時に JPEG で発生するブロック・ノイズやモスキート・ノイズが目立ちにくいという利点がある。

複数コンポーネントの圧縮 RGB, YUV などの複数コンポーネントから構成されるカラー画像を圧縮できる。1 コンポーネントのビット精度も、すべてのアプリケーションをカバーするのに十分な精度を持っている。

表 2-1: 8×8 量子化テーブルの例 (文献⁽⁹⁾引用)

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

可逆・非可逆圧縮 可逆型ウェーブレットフィルタ (非量子化) を用いることで、可逆圧縮が実現できる。また、同一の手法で非可逆圧縮も可能であり、編集やビットレート変換などの際に効果的である。

ランダムアクセス 画面中の特定領域のみを重視して圧縮率を高く設定するなどの ROI (Region Of Interest) が実現できる。デコーダ側では ROI の部分から最初に復号することが可能である。医療用画像のような、非常に解像度が高く、画像全体を一律の圧縮率で圧縮するとデータが大きくなる場合、特に重要な患部だけを ROI で符号化する応用が考えられる。

スケーラビリティ サブバンド分割による空間スケーラビリティ、ポスト量子化による品質スケーラビリティなど高いスケーラビリティを持ち、ワンソースマルチユースを実現する。

JPEG 2000 は JPEG でターゲットとしているカラー静止画像に加え、医療用画像や衛星写真画像のような、超高解像度画像もターゲットにしている。これらの画像は数千画素 $\times$ 数千ラインからなる空間解像度を持ち、カラー静止画像で一般的な 8 [bit] よりも大きなビット深度を持っている。一方で、JPEG では 8 [bit] のビット深度までしか対応していない。また、デジタルシネマは JPEG 2000 の符号化方式を動画画像に拡張した Motion JPEG 2000 により符号化されている。Motion JPEG 2000 は MPEG (Moving Picture Experts Group) のようにフレーム間の相関を利用する動き補償予測を行わず、フレーム単位で独立に符号化する。そのため、符号化効率は MPEG よりも劣るが、符号化ストリームに対する編集の容易さやランダムアクセスの容易さの点で優れている。

現在では、JPEG 2000 はこのような特殊なアプリケーションにおいて用いられるに留まっており、一般的なカラー静止画像については JPEG が依然として利用されている。その原因として、近年の HDD や光ディスクドライブなどのストレージの急激な価格低下や ADSL や光ファイバなどの高速インターネット回線の普及により、静止画像の圧縮率がそれほど求

められなくなったことがある。デジタルスチルカメラでの JPEG 2000 サポートも期待されるが、専用チップが必要な上、演算コストの増加分に見合うだけの符号化効率が得られないなどの難しい側面もある。また、JPEG と異なり、パーソナルコンピュータのオペレーティングシステムのトップシェアを誇る Microsoft Windows で標準サポートされていないことも普及しない要因である。さらに、2007 年 11 月に Microsoft が独自に開発した HD Photo と呼ばれる静止画像ファイルフォーマットが JPEG XR (eXtension Range) として新たに標準化されることが決定した⁽¹¹⁾。このため、JPEG 2000 はカラー静止画像の圧縮方式として普及せずに終わる可能性がある。

JPEG 2000 の技術

JPEG 2000 の基本構造も JPEG と同様に図 2-1 である。周波数変換として、JPEG が DCT を用いるのに対して、JPEG 2000 では DWT を用いる。DCT は周期関数を用いた直交変換であるため、定常信号を表現するのに適しているが、非定常信号には不向きである。一方で、DWT はウェーブレットとよばれる局在波を用いた変換であるため、非定常信号を表現するのに適している。一般的な自然画像は局所的な変化を多く含む非定常信号であるため、DWT は DCT より適切に自然画像を表現することができる。DWT は本研究と深く関連する技術であるため、第 2.3 節において詳細を述べる。JPEG 2000 では、変換係数の量子化にスカラー量子化とポスト量子化が用意されている。また、エントロピー符号化は EBCOT (Embedded Block Coding with Optimized Truncation) と呼ばれる技術が用いられている。JPEG 2000 の符号化アルゴリズムを図 2-4 に示す。それぞれの処理について以下に述べる⁽¹²⁾⁽¹³⁾⁽¹⁴⁾。

DC レベル変換

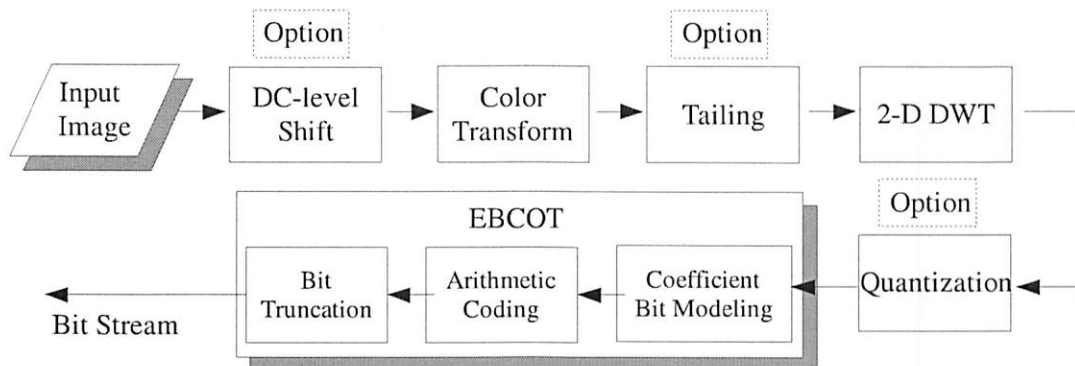
入力信号が正の値である場合には、各信号値から信号のダイナミックレンジの半分を減算するレベルシフトを行う。逆に復号時には各信号値に信号のダイナミックレンジの半分を加算するレベルシフトを行う。なお、DC レベル変換はオプションである。

コンポーネント変換

入力画像が RGB の色空間で構成される場合に、YCbCr と呼ばれる輝度成分と色差成分からなる色空間に線形変換される。色変換には可逆変換の RCT (Reversible multiple Component Transformation) と非可逆変換の ICT (Inreversible multiple Component Transformation) の 2 つの変換が定義されている。

タイル分割

JPEG 2000 では、画像をタイルと呼ばれるいくつかの矩形領域に分割し、以後、タイルを 1 枚の独立した画像として扱う処理が可能である。分割された各タイルは、タイル境界

図 2-4: JPEG 2000 符号化アルゴリズム (文献⁽¹²⁾ 参照)

を越えて画素参照されることはなく、それぞれ独立に符号化される。タイルは、符号化処理システムのメモリ使用量のサイズ依存性を低減するとともに、並列に処理することによって処理の高速化を図ることも可能となり、サイズの大きい画像を符号化する際に特に有効である。なお、タイル分割はオプションである。

離散ウェーブレット変換

DWT については第 2.3 節において詳細を述べる。

量子化

係数のダイナミックレンジを削減する。JPEG 2000 では、スカラー量子化とポスト量子化の 2 つの量子化が定義されている。スカラー量子化には、各解像度レベルにおける全サブバンドの量子化を規定する方式と、最下位の解像度レベルの LL サブバンドのみを規定して、残りのサブバンドは既存の定義式を用いて規定する方式の 2 種類がある。前者を明示的な量子化 (explicit quantization)、後者を暗黙的な量子化 (implicit quantization) と呼ぶ。なお、可逆変換時には量子化は行わない。一方、ポスト量子化では、完成した符号列の下位ビットプレーンを切り捨てることによって量子化を行う。

EBCOT

JPEG 2000 では、エントロピー符号化として EBCOT と呼ばれるブロックベースのビットプレーン符号化が用いられる。EBCOT 処理の段階を以下に示す。

- コードブロック分割
- 係数ビットモデリング
- 算術符号化

コードブロック分割 入力されたサブバンド係数を、同一サイズのコードブロックと呼ばれる矩形領域に分割する。EBCOT では、このコードブロックを処理単位として、以降の処理を独立に行う。コードブロックのサイズは、面積が16 から 4096 の範囲で自由に決めることが可能である。

係数ビットモデリング 各コードブロックの変換係数をビットプレーン分解した後、それぞれのビットプレーン内の各係数ビットを判定し、significance propagation pass (有意である係数が周囲にある有意でない係数の符号化パス), magnitude refinement pass (有意である係数の符号化パス), cleanup pass (残りの係数の符号化パス) の3つの符号パスに分解する。

算術符号化 ビットプレーン分解された各サブバンド係数は、MQ コーダと呼ばれる符号器によってエントロピー符号化され、圧縮されたビットストリームが生成される。MQ コーダは、2 値画像の標準規格である JBIG2 (Joint Bitonal Image Group) で規定された、学習型の 2 値算術符号化である。

2.3 離散ウェーブレット変換

2.3.1 連続ウェーブレット変換

ウェーブレットは局在する不規則な振動波をあらわす関数の総称であり、ウェーブレットを用いた変換はウェーブレット変換 (WT: Wavelet Transform) と呼ばれる。WT は FT と異なり、時間と周波数を同時に解析できる。音楽、スピーチ、画像などは局所的に信号の性質が変化する非定常的な信号であるため、このような信号の解析に時間方向の解析が可能な WT は適している⁽¹⁵⁾。

連続ウェーブレット変換 (CoWT: Continuous Wavelet Transform) ではマザーウェーブレットと呼ばれる関数 $\psi(t) \in L^2(\mathbb{R})$ を定義する。現在では多くのマザーウェーブレットが提案されており、Haar ウェーブレット, Gabor ウェーブレット, Daubechies ウェーブレットなどが良く知られている。マザーウェーブレット ψ を平行移動 (シフト) および伸縮 (スケーリング) することで得られる関数群 $\psi_{a,b}(t)$ を式 (2.3) により定義する。

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad \text{where } a, b \in \mathbb{R} (a > 0) \quad (2.3)$$

ここで、 a はスケーリングパラメータと呼ばれマザーウェーブレットの伸縮比を制御し、 b はシフトパラメータと呼ばれマザーウェーブレットの時間方向のシフト量を制御する。また、 $\|\psi_{a,b}(t)\|_2 = \|\psi(t)\|_2$ となるよう正規化されている。マザーウェーブレットは次のアドミッシブルの条件を満たす必要がある。

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty \quad (2.4)$$

ここで、 $\Psi(\omega)$ は $\psi(t)$ の FT である。 $\Psi(\omega)$ は遠方において減衰することが一般的で、このとき式 (2.4) が成立するためには $\omega = 0$ において $\Psi(0) = 0$ とならなければならない。これ

は時間領域において $\psi(t)$ の積分が 0 となることと等価である。すなわち、アドミッシブルの条件より式 (2.5) が導かれる。

$$\int_{-\infty}^{\infty} \psi(t) dt = \Psi(0) = 0 \quad (2.5)$$

これはウェーブレットがバンドパスフィルタの役割をすることを示している。

アドミッシブルの条件を満たすウェーブレットを用いて、関数 $f(t) \in \mathbb{R}$ の CoWT は式 (2.6) により定義される。

$$\text{CoWT}_f(a, b) = \int_{-\infty}^{\infty} \overline{\psi_{a,b}(t)} f(t) dt = \langle \psi_{a,b}(t) | f(t) \rangle \quad (2.6)$$

ここで、 $\bar{\psi}$ は ψ の共役複素数である。また、 $\langle f | g \rangle$ は 2 関数 $f, g \in L^2(\mathbb{R})$ の内積をあらわし式 (2.7) である。

$$\langle f | g \rangle = \int_{-\infty}^{\infty} \overline{f(x)} g(x) dx \quad (2.7)$$

式 (2.6) は各パラメータ a, b における $f(t)$ と $\psi_{a,b}$ の相関値を計算していると解釈できる⁽¹⁷⁾。パラメータ $\hat{a}$ を固定し、パラメータ b を変化させたときの $\text{CoWT}_f(\hat{a}, b)$ は、時間-周波数分解能を一定のまま異なる時間位置を解析した結果である。また、パラメータ $\hat{b}$ を固定し、パラメータ a を増加させたときの $\text{CoWT}_f(a, \hat{b})$ は、位置 $\hat{b}$ における $f(t)$ の大域的な性質をあらわし、減少させたときの $\text{CoWT}_f(a, \hat{b})$ は $f(t)$ の局所的な性質をあらわす。

式 (2.6) に対応する逆変換が存在し、ウェーブレット係数 $\text{CoWT}_f(a, b)$ から原信号 $f(t)$ を再構成することができる。

$$f(t) = \frac{1}{C_\psi} \iint_{-\infty}^{\infty} \text{CoWT}_f(a, b) \psi_{a,b}(t) \frac{da db}{a^2} \quad (2.8)$$

2.3.2 パラメータの離散化

CoWT は連続的に変化するパラメータ (a, b) に基づきウェーブレット係数 $\text{CoWT}_f(a, b)$ が求まる。このとき、不確定性原理 (uncertainty principle)⁽¹⁶⁾ より時間-周波数分解能の面積は有限値より小さくすることができない。このため、2 つのパラメータの組 (a', b') 、 (a, b) があまりにも近接していると、時間-周波数分解能の領域が重複し、ウェーブレット係数 $\text{CoWT}_f(a', b')$ と $\text{CoWT}_f(a, b)$ の間に冗長性が発生する。この様子を図 2-5 に示す。図 2-5 において $2\Delta_t$ は時間幅、 $2\Delta_f$ は周波数幅をあらわしている。このとき、 $2\Delta_t \times 2\Delta_f$ により作られる長方領域が時間-周波数分解能の解像度をあらわしており、その面積は有限値より小さくすることはできない。よって、2 つのパラメータ (a, b) の値が近いとき、領域の重複が発生する。このような領域の重複を避けるためには、式 (2.9) のようにパラメータ (a, b) を離散化する必要がある。

$$a = a_0^j \quad \text{and} \quad b = k a_0^j b_0 \quad \text{where} \quad j, k \in \mathbb{Z} \quad (2.9)$$

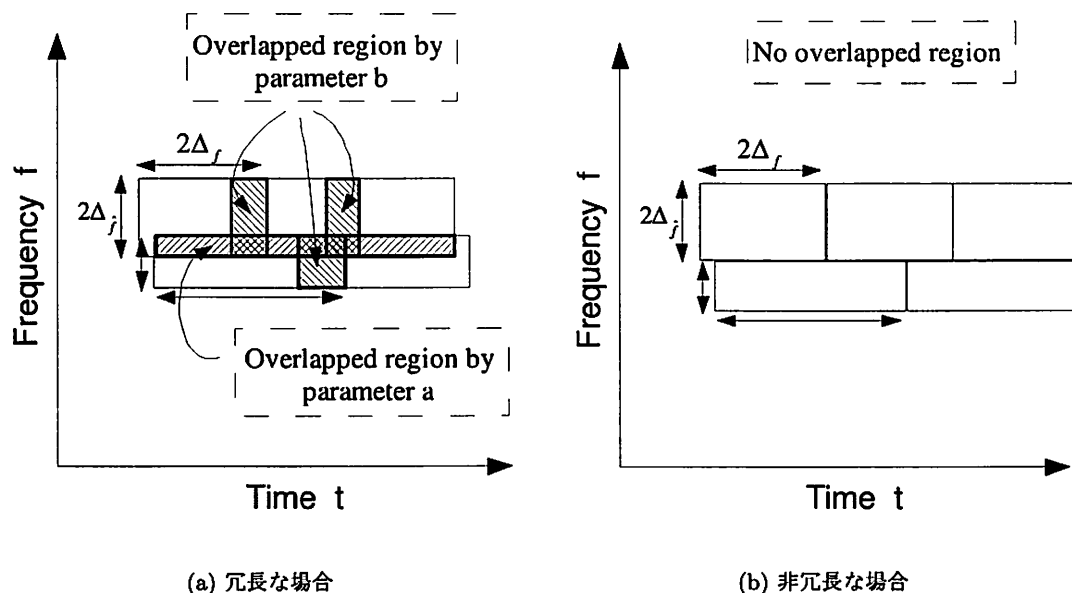


図 2-5: 時間-周波数表現における領域の重複と冗長性

パラメータの離散化により，関数群 $\psi_{a,b}(t)$ は関数群 $\psi_{j,k}(t)$ となる．

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a_0^{-j}t - kb_0) \quad (2.10)$$

$\psi_{j,k}(t)$ を用いたウェーブレット変換は DWT と呼ばれる．式 (2.6) の CoWT に対応して，DWT では式 (2.11) より離散ウェーブレット係数 $D_f(j, k)$ が求まる．

$$D_f(j, k) = \int_{-\infty}^{\infty} \overline{\psi_{j,k}(t)} f(t) dt = \langle \psi_{j,k}(t) | f(t) \rangle \quad (2.11)$$

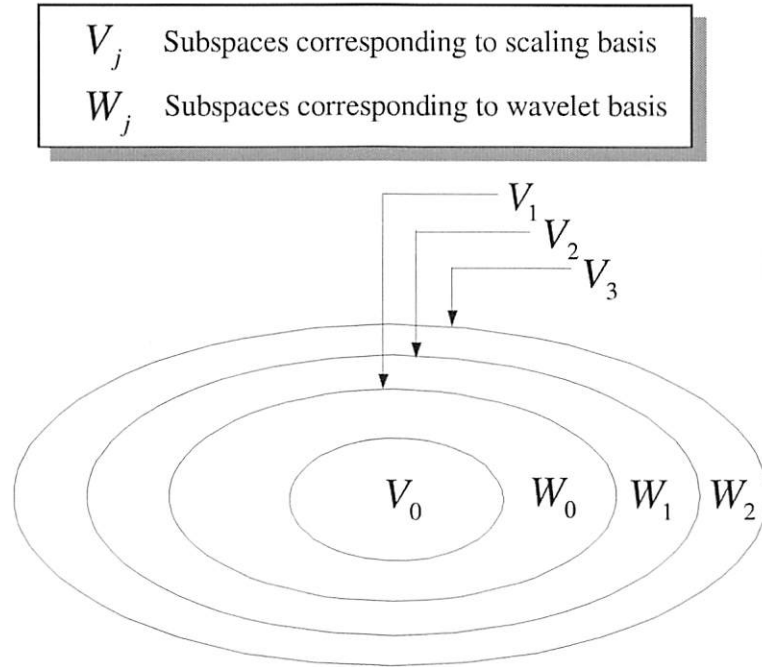
また，式 (2.8) の CoWT 逆変換に対応して，DWT では式 (2.12) の逆変換により離散ウェーブレット係数 $D_f(j, k)$ から原信号 $f(t)$ を再構成することができる．

$$f(t) = \sum_j \sum_k D_f(j, k) \psi_{j,k}(t) \quad (2.12)$$

一般的に広く利用される DWT は， $a_0 = 2^{-1}$ ， $b_0 = 1$ である．また，DWT では離散ウェーブレット係数から元の信号を完全に再構成するためにマザーウェーブレットは基底関数でなければならない．このため，Gabor ウェーブレットのように連続系では利用できるが離散系では利用が難しいマザーウェーブレットが存在する⁽¹⁵⁾．

2.3.3 多重解像度解析の原理

次の条件を満たす $L^2(\mathbb{R})$ の閉部分空間の列 $\{V_j\}_{j \in \mathbb{Z}}$ を多重解像度解析 (MRA : Multi-Resolution Analysis) と呼ぶ．



$$V_1 = V_0 \oplus W_0$$

$$V_2 = V_1 \oplus W_1 = V_0 \oplus W_1 \oplus W_2$$

$$V_3 = V_2 \oplus W_2 = V_0 \oplus W_1 \oplus W_2 \oplus W_2$$

図 2-6: 多重解像度解析における空間の入れ子構造

1. $V_j \subset V_{j+1}, j \in \mathbb{Z}$
2. $V_{-\infty} = \{0\}$ and $V_{\infty} = L^2$
3. $f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}$
4. $f(t) \in V_0 \Leftrightarrow f(t-k) \in V_0, \forall k \in \mathbb{Z}$
5. for $\varphi(t) \in V_0, \{\varphi(t-k)\}_{k \in \mathbb{Z}}$ is Orthogonal basis of the subspace V_0

条件 (5) の関数 $\varphi(t)$ はスケーリング関数と呼ばれ、正規直交基底の代わりに Riesz 基底を用いて条件を緩める場合もある。条件 (1) は空間が入れ子構造になっていることをあらわす。条件 (3) は条件 (1) の全空間がスケール変換で構成されていることをあらわす。条件 (4) は平行移動に対して空間が普遍であることをあらわす。

スケーリング関数を用いた関数群 $\varphi_{j,k}(t)$ は式 (2.13) であらわされ、それぞれ空間 V_j の直交基底となる。

$$\varphi_{j,k}(t) = 2^{j/2} \varphi(2^j t - k), \quad \text{where } j, k \in \mathbb{Z} \quad \text{and} \quad \int \varphi(t) dt = 1 \quad (2.13)$$

条件 (1) の空間の入れ子構造より, Hilbert の直交分解定理から V_{j+1} は V_j とその直交補空間 W_j に分解することができる (図 2-6).

$$V_{j+1} = V_j \oplus W_j, \quad V_j \perp W_j \quad (2.14)$$

このとき, 式 (2.15) からなる関数群 $\psi_{j,k}(t)$ がそれぞれ空間 W_j の直交基底となるようなマザーウェーブレット $\psi(t)$ が存在する.

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad \text{where } j, k \in \mathbb{Z} \quad (2.15)$$

また, 条件 (1) の空間の入れ子構造と条件 (3) の空間のスケール関係より, スケーリング関数には式 (2.16) のツースケール関係が成り立つ.

$$\varphi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_0[n] \varphi(2t - n), \quad \text{where } n \in \mathbb{Z} \quad (2.16)$$

ここで, アドミッシブルの条件より $\sum_n h_0[n] = \sqrt{2}$ である. ウェーブレットについても同様の式 (2.17) が成り立つ.

$$\psi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h_1[n] \varphi(2t - n), \quad \text{where } n \in \mathbb{Z} \quad (2.17)$$

ここで, $\sum_n h_1[n] = 0$, $h_1[n] = (-1)^n h_0[-n+1]$ である. $h_0[n]$ と $h_1[n]$ はツースケール数列と呼ばれ, それぞれローパスフィルタとハイパスフィルタの係数とみなすことができる. ウェーブレット係数は $f(t)$ の空間 V_{j+1} への射影から, 1 レベル低解像度の空間 V_j への射影の際に生じる情報の損失をあらわしている. 多重解像度解析を用いると式 (2.12) の DWT の逆変換は, 式 (2.13) と式 (2.15) で定義したスケーリング関数とウェーブレットを用いて式 (2.18) であらわせる.

$$f(t) = \sum_{k=-\infty}^{\infty} C_f(J_0, k) \varphi_{J_0, k}(t) + \sum_{j=J_0}^{J_1} \sum_{k=-\infty}^{\infty} D_f(j, k) \psi_{j, k}(t) \quad (2.18)$$

ここで, $C_f(J_0, k)$ はスケーリング係数と呼ばれ, 式 (2.19) により求まる.

$$C_f(J_0, k) = \int_{-\infty}^{\infty} \overline{\varphi_{J_0, k}(t)} f(t) dt = \langle \varphi_{J_0, k}(t) | f(t) \rangle \quad (2.19)$$

J_0 は最低解像度, J_1 は最高解像度であり $J_1 > J_0$ である.

2.3.4 フィルタバンクによる実装

$C_f(j, k)$ と $D_f(j, k)$ は, 信号 $f(t)$ を空間 V_j と W_j に射影した際に発生するスケーリング係数とウェーブレット係数であるとする. 式 (2.16) と式 (2.17) のようなスケーリング関数とウェーブレットのスケール関係より, 連続する低解像度のスケーリング係数 $C_f(j-1, k)$

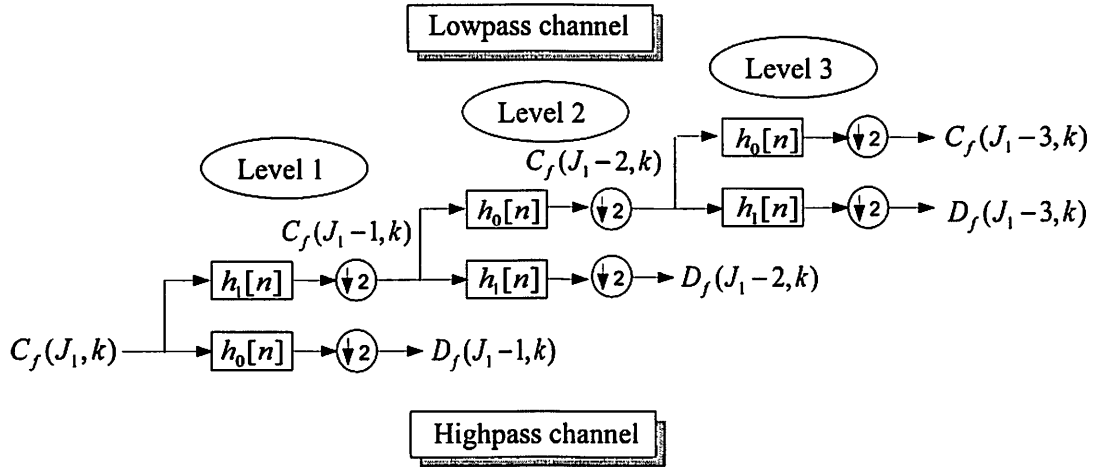


図 2-7: 1次元離散ウェーブレット変換における分割フィルタバンク

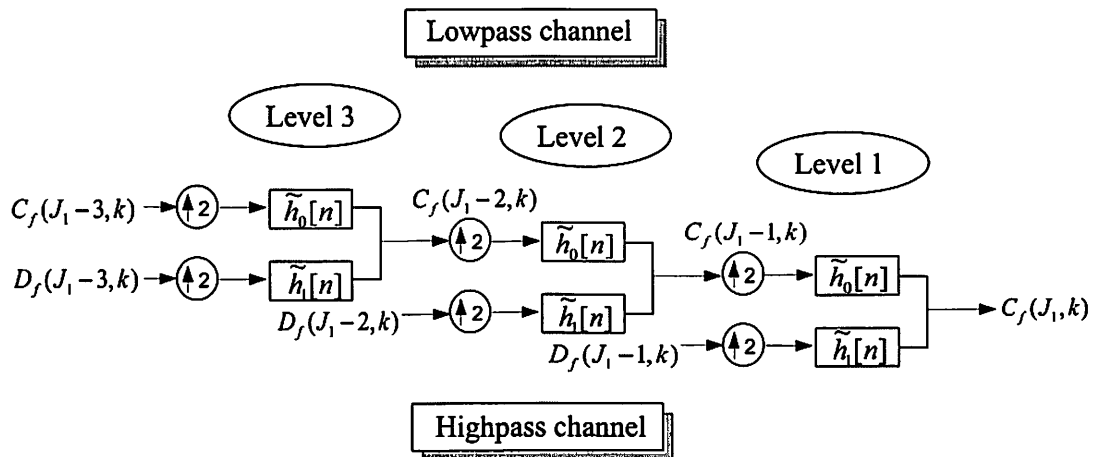


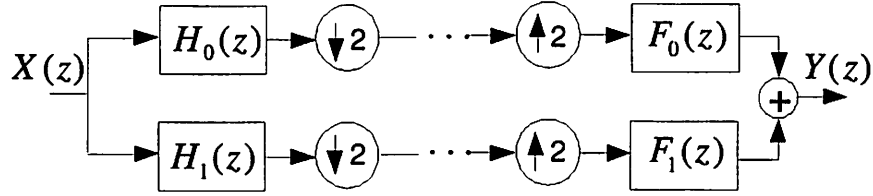
図 2-8: 1次元離散ウェーブレット変換における合成フィルタバンク

はトウスケール数列 $h_0[n]$ と高解像度のスケーリング係数 $C_f(j, k)$ の線形和により再帰的に求めることができる。ウェーブレット係数に関しても同様である。

$$\begin{aligned}
 C_f(j-1, k) &= \sum_n h_0[n-2k] C_f(j, n) \\
 D_f(j-1, k) &= \sum_n h_1[n-2k] C_f(j, n)
 \end{aligned} \tag{2.20}$$

これは図 2-7 と図 2-8 のような木構造のフィルタバンクとして実現することができる。図 2-7 は分割フィルタバンクと呼ばれ、信号をスケーリング係数とウェーブレット係数へ分解する処理である。図 2-8 は合成フィルタバンクと呼ばれ、スケーリング係数とウェーブレット係数から信号を再構成する処理である。

図 2-7 において $\downarrow 2$ はダウンサンプリングであり信号を間引く処理である。 $h_0[n]$ はローパスフィルタ、 $h_1[n]$ はハイパスフィルタである。分割フィルタバンクの木構造は、入力信号

図 2-9: 2 分割フィルタバンク (文献⁽¹⁸⁾ 参照)

をレベル J_1 におけるスケーリング係数列 $C_f(J_1, k)$ とし、ローパスフィルタとハイパスフィルタを階層的に適用する構造である。入力信号の詳細部分 (高周波) はレベル j のウェーブレット係数 $D_f(j, k)$ で階層的に表現される。一般的な DWT では各レベルにおいて、ダウンサンプリングによりスケーリング係数とウェーブレット係数を半分に間引く。よって DWT は変換前と変換後においてデータ数が変化しない非冗長な変換である。

図 2-8 において $\uparrow 2$ はアップサンプリングであり信号を零補間する処理である。 $\tilde{h}_0[n]$ はローパスフィルタ、 $\tilde{h}_1[n]$ はハイパスフィルタである。合成フィルタバンクは図 2-7 の分割処理に対する逆処理であり、高解像度のスケーリング係数を連続する低解像度のスケーリング係数とウェーブレット係数から式 (2.21) により再構成する。

$$C_f(j+1, k) = \sum_n \tilde{h}_0[k-2n]C_f(j, n) + \sum_n \tilde{h}_1[k-2n]D_f(j, n) \quad (2.21)$$

2.3.5 完全再構成

サブバンド符号化に用いるフィルタバンクは完全再構成 (PR: Perfect Reconstruction) を満たす必要がある⁽¹⁸⁾。完全再構成とはフィルタバンクによる分割・合成処理後の再構成信号が入力信号と完全に一致する性質である。完全再構成では図 2-9 において $X(z) = Y(z)$ が成立する。図 2-9 は Z 変換を用いた表現をしており、 $H_0(z), H_1(z)$ はそれぞれ $h_0[n], h_1[n]$ の Z 変換、 $F_0(z), F_1(z)$ はそれぞれ $\tilde{h}_0[n], \tilde{h}_1[n]$ の Z 変換である。このとき入力信号 $X(z)$ と出力信号 $Y(z)$ の関係は式 (2.22) となる。

$$Y(z) = \frac{1}{2}\{H_0(z)F_0(z) + H_1(z)F_1(z)\}X(z) + \frac{1}{2}\{H_0(-z)F_0(z) + H_1(-z)F_1(z)\}X(-z) \quad (2.22)$$

式 (2.22) において右辺第二項はダウンサンプリングおよびアップサンプリングにより生じたエイリアシング成分とイメージング成分であり、 $Y(z) = X(z)$ となるためにはこの項が 0 とならなければならない。すなわち、式 (2.23) と式 (2.24) が成り立つ。

$$H_0(-z)F_0(z) + H_1(-z)F_1(z) = 0 \quad (2.23)$$

$$H_0(z)F_0(z) + H_1(z)F_1(z) = 2 \quad (2.24)$$

式 (2.23) よりフィルタの相互関係を式 (2.25) で規定すると、エイリアシングとイメージング成分の除去が可能となる。

$$\begin{aligned} F_0(z) &= H_1(-z) \\ F_1(z) &= -H_0(-z) \end{aligned} \quad (2.25)$$

式 (2.25) の条件の下で式 (2.24) を表現すると次のようになる。

$$H_0(z)H_1(-z) - H_0(-z)H_1(z) = 2 \quad (2.26)$$

よって完全再構成フィルタバンクを構成するためには、式 (2.26) の関係を満たすアナライザフィルタ $H_0(z)$, $H_1(z)$ が必要である。直交ウェーブレットを構成するフィルタを作るためには式 (2.26) の条件下で、さらにアナライザフィルタ $H_0(z)$, $H_1(z)$ に直交関係が必要となる。フィルタには有限タップの FIR (Finite Impulse Response) フィルタと無限タップの IIR (Infinite Impulse Response) フィルタがある。また、信号処理に用いるフィルタは線形位相をもつことが望ましい。完全再構成、直交条件、線形位相の 3 条件を満足するフィルタバンクは IIR フィルタを用いて実現できる。しかし、ウェーブレットは有限時間で減衰するコンパクトサポート特性をもつ関数が望ましい。これを実現できるのは FIR フィルタである。FIR フィルタでは、完全再構成、直交条件、線形位相の 3 条件を満足するフィルタは Haar フィルタのみであることが知られている。しかし、Haar フィルタは矩形関数であるため滑らかな信号の表現には適していない。そこで、滑らかな関数をもつ FIR フィルタにおいて線形位相を維持したい場合、直交関係を双直交関係へと緩和する必要がある。このような FIR フィルタにより実現されるウェーブレットは双直交ウェーブレットと呼ばれる。双直交ウェーブレットではアナライジングウェーブレット $\tilde{\psi}(t)$ とシンセサイジングウェーブレット $\psi(t)$ のペア $\{\tilde{\psi}(t), \psi(t)\}$ と、スケーリング関数のペア $\{\tilde{\varphi}(t), \varphi(t)\}$ が分割処理と合成処理の双対基底となっている。このとき式 (2.16) と式 (2.17) に対応するツースケール関係は式 (2.27) となる。

$$\begin{aligned} \varphi(t) &= \sqrt{2} \sum_n h_0[n] \varphi(2t - n), \quad \tilde{\varphi}(t) = \sqrt{2} \sum_n \tilde{h}_0[n] \tilde{\varphi}(2t - n) \\ \psi(t) &= \sqrt{2} \sum_n h_1[n] \varphi(2t - n), \quad \tilde{\psi}(t) = \sqrt{2} \sum_n \tilde{h}_1[n] \tilde{\varphi}(2t - n) \end{aligned} \quad (2.27)$$

また、ウェーブレットとスケーリング関数には式 (2.28) の直交関係が成立する。

$$\begin{aligned} \langle \psi_{j,k} | \tilde{\psi}_{l,m} \rangle &= \delta_{j,l} \delta_{k,m} \\ \langle \varphi_{j,k} | \tilde{\varphi}_{l,m} \rangle &= \delta_{j,l} \delta_{k,m} \end{aligned} \quad (2.28)$$

ここで $\delta_{i,j}$ はクロネッカーのデルタであり式 (2.29) で与えられる。

$$\delta_{i,j} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (2.29)$$

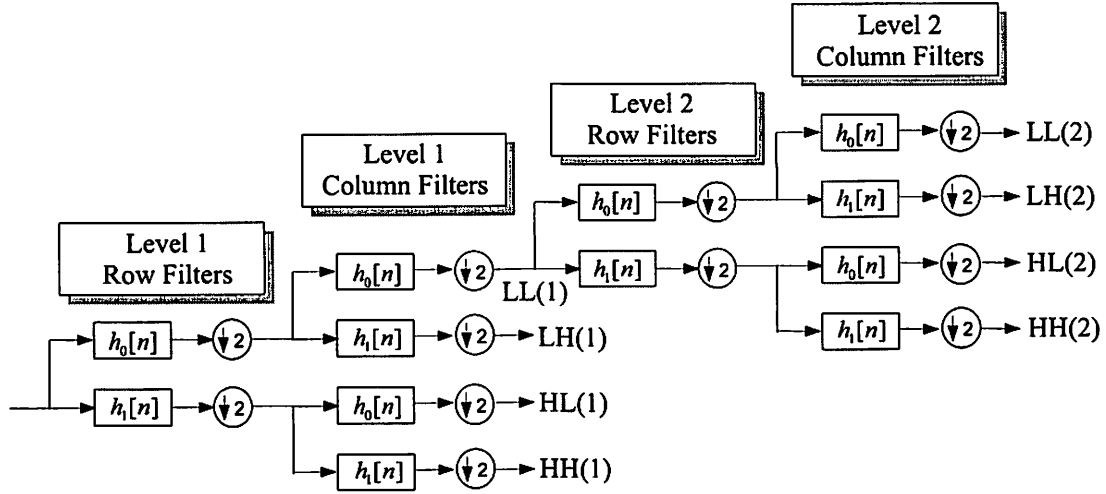


図 2-10: 2次元離散ウェーブレット変換におけるフィルタバンク

式 (2-18) に対応する双直交ウェーブレットにおける DWT の逆変換は式 (2-30) となる。

$$f(t) = \sum_{k=-\infty}^{k=\infty} \langle \varphi_{J_0,k}(t) | f(t) \rangle \tilde{\varphi}_{J_0,k}(t) + \sum_{j=J_0}^{J_1} \sum_{k=-\infty}^{\infty} \langle \psi_{j,k}(t) | f(t) \rangle \tilde{\psi}_{j,k}(t) \quad (2.30)$$

JPEG 2000 では, Daubechies 9/7 タップウェーブレットと整数型 5/3 タップウェーブレットが用いられているが, これらはいずれも双直交ウェーブレットである。

2.3.6 2次元離散ウェーブレット変換

図 2-7 の分割フィルタバンクを用いた DWT は 1 次元信号を対象としたものであった。一方で, 本研究の対象である画像信号は 2 次元信号である。そこで, 1 次元 DWT (1D DWT) を 2 次元へ拡張する必要がある。一般的な 2 次元 DWT (2D DWT) では分離型フィルタバンクが用いられる。分離型フィルタバンクの構造は図 2-10 のように 1 次元のフィルタバンクを横方向, 縦方向の順に適用する。この構造は入力画像の水平方向 (0°) を帯域分割する HL サブバンド, 垂直方向 (90°) を帯域分割する LH サブバンド, 対角方向 ($\pm 45^\circ$) を帯域分割する HH サブバンドの 3 種類の高域サブバンドと低周波成分の LL サブバンドに分解する。このうち LL サブバンドに対して, 同様の処理を階層的に行うことで画像信号の多重解像度解析が可能となる。2D DWT は 1D DWT と同様に各レベルの横方向, 縦方向の処理においてダウンサンプリングがなされるため, 入力画像とデータ数が増えない非冗長な変換である (図 2-11)。

自然画像に対してレベル 2 までの 2D DWT を行った様子を図 2-12 に示す。図 2-12(a) は 512×512 画素サイズのグレースケール標準画像 Lena, 図 2-12(c) は 512×512 画素サイズのグレースケール標準画像 Barbara, 図 2-12(e) は 512×512 画素サイズのグレースケール標準画像 Goldhill である。また, 図 2-12(b) は Lena に対する 2D DWT のウェーブレット

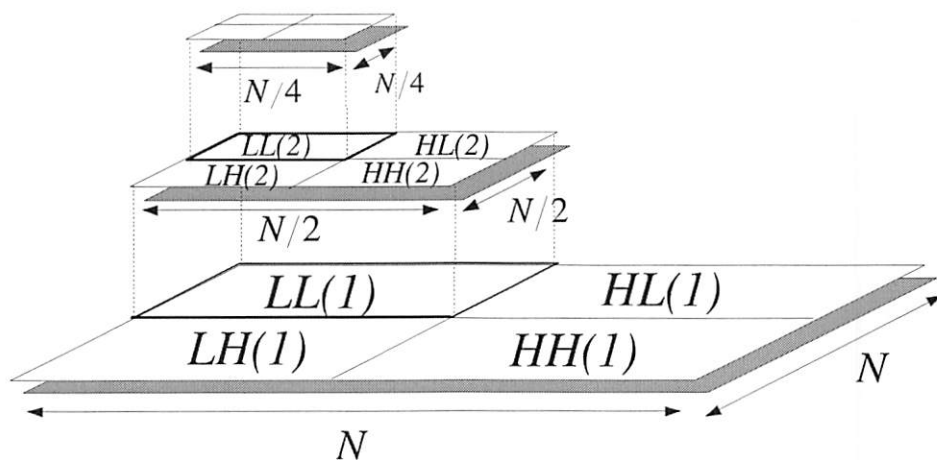


図 2-11: 2次元離散ウェーブレット変換の階層構造

表 2-2: Daubechies 9/7 タップウェーブレットのフィルタ係数

$h_0[n]$	$h_1[n]$
0.0378	
-0.0238	-0.0645
-0.1106	0.0407
0.3774	0.4181
0.8527	-0.7885
0.3774	0.4181
-0.1106	0.0407
-0.0238	-0.0645
0.04	

ト係数, 図 2-12(d) は Barbara に対する 2D DWT のウェーブレット係数, 図 2-12(f) は Goldhill に対する 2D DWT のウェーブレット係数である. ここでは離散ウェーブレットとして表 2-2 に示す Daubechies 9/7 タップウェーブレットを用いた.

2.4 離散ウェーブレット変換の問題点

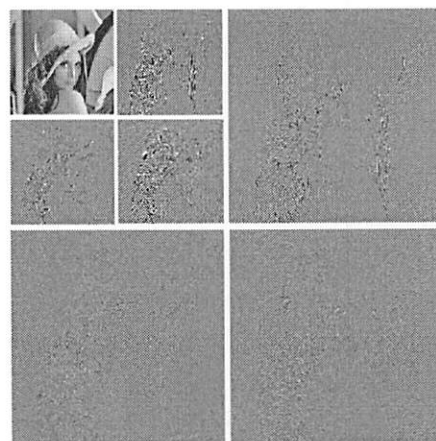
2.4.1 シフト不変性の欠如

DWT はシフト不変性が成立しない⁽¹⁹⁾⁽²⁰⁾. シフト不変性とは, 時間軸上における入力信号の微小シフトに対して, 変換係数が一定の変化をする特性であり, 式 (2.31) のように表現される.

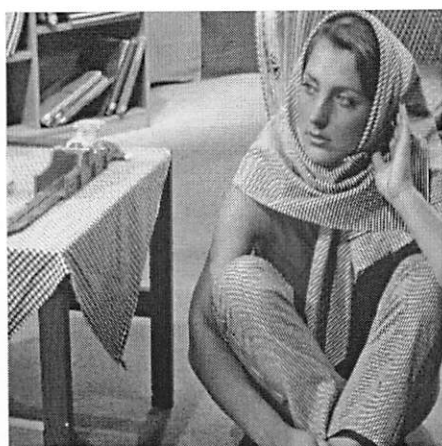
$$y(n) = T[x(n)] \rightarrow y(n-k) = T[x(n-k)] \quad (2.31)$$



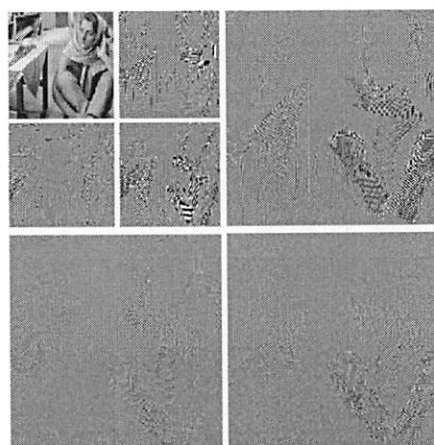
(a) 原画像 Lena



(b) Lena に対する 2D DWT



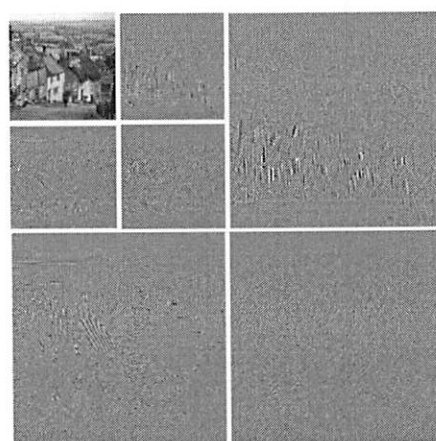
(c) 原画像 Barbara



(d) Barbara に対する 2D DWT



(e) 原画像 Goldhill



(f) Goldhill に対する 2D DWT

図 2-12: 自然画像に対する 2 次元離散ウェーブレット変換

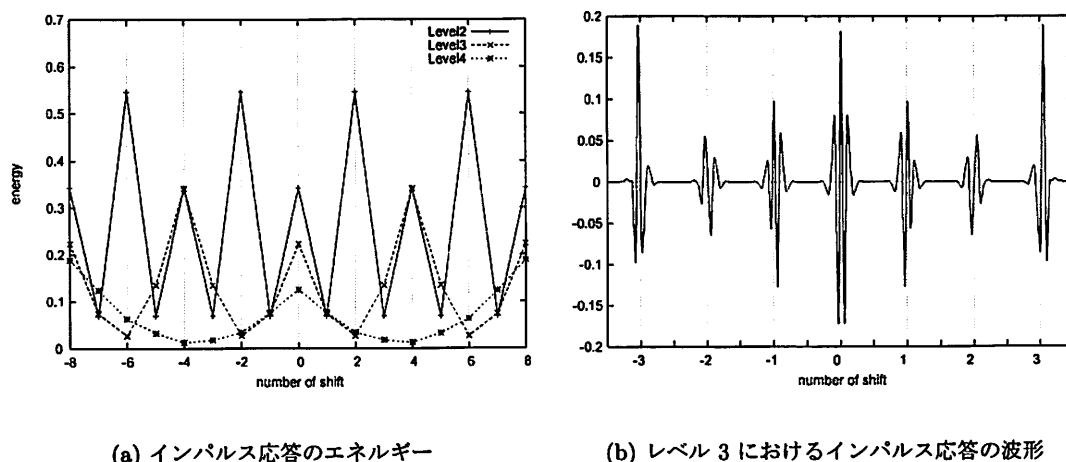


図 2-13: 離散ウェーブレット変換におけるシフト不変性の欠如

これが欠如している場合、入力信号のシフトに対して変換係数の値が予測できないほど変動する。よって変換係数から入力信号のシフトを追跡することができない。この様子を図 2-13 に示す。図 2-13 はインパルス信号をシフトさせながら入力させた場合の DWT 係数のインパルス応答をあらわしている。図 2-13(a) は各レベルにおける DWT 係数のインパルス応答のエネルギー、図 2-13(b) はレベル 3 における実際のインパルス応答の波形である。インパルス信号をわずかに数サンプルシフトさせただけでもエネルギーは激しく振動し、波形も不規則に変形していることがわかる。このためシフト不変性を必要とするアプリケーションにおいて DWT を用いることは難しい。

DWT においてシフト不変性が成立しない原因は、ダウンサンプリングに伴うエイリアシングの発生である⁽¹⁹⁾。第 2.3.5 節の再構成条件において、式 (2.22) が第一項のみから構成される場合 (すなわち、第二項 = 0)、 $Y(z)$ と $X(z)$ は単一の線形シフト不変の Z 変換で結ばれる。よって、エイリアシングが発生しない場合はシフト不変性が成立する。しかし、実際はダウンサンプリングにより発生するエイリアシング成分の第二項が存在する。式 (2.22) は図 2-9 のレベル 1 における 2 分割フィルタバンクに対応しているため、エイリアシング成分の項が 1 つしかない。しかし、レベル 2 以降についてもレベル j に対応するサンプリングレートを $\downarrow 2^j$ として、エイリアシング成分の項が多数生成される。DWT ではこれらの項すべてを 0 とすることができない。よって、エイリアシング成分の項が残り各レベルにおいてシフト不変性が成り立たない。一方で、第 3 章で述べる DTCWT では、各レベルでエイリアシング成分の項を近似的に 0 とすることで近似的なシフト不変性を実現している。

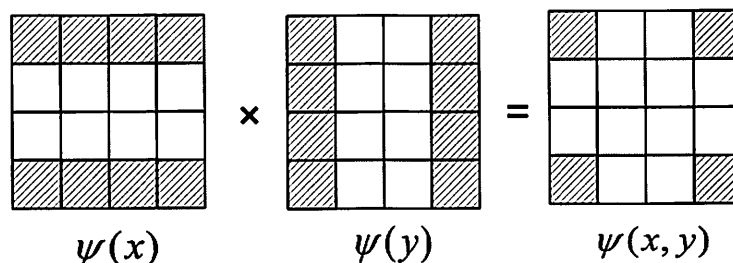
2.4.2 低性能な方向分離特性

第 2.3.6 節において 2D DWT は分離型フィルタを用いて 1D DWT を拡張することで実現されることを述べた。ここで、1 次元におけるウェーブレットを $\psi(x)$ 、スケーリング関数

を $\varphi(x)$ とすると、2次元における各サブバンド (HL, LH, HH) のウェーブレットは式 (2.32) により与えられる。

$$\begin{aligned} \text{HL} : \psi^V(x, y) &= \varphi(x)\psi(y) \\ \text{LH} : \psi^H(x, y) &= \psi(x)\varphi(y) \\ \text{HH} : \psi^D(x, y) &= \psi(x)\psi(y) \end{aligned} \quad (2.32)$$

例として $\psi^D(x, y)$ の周波数スペクトルの生成モデルを以下に示す。



DWT では φ と ψ が (複素関数ではなく) 実関数であるため、1次元変換の際に周波数スペクトルが対称に発生する。2次元変換ではこの対称なスペクトルが水平方向および垂直方向で発生し、それらが乗算される。よって、2次元周波数スペクトルは四隅で発生する。 $\psi^V(x, y)$, $\psi^H(x, y)$ についても同様にして、結果的に図 2-14(a) のような 2次元帯域分割がなされる。また、全方向のエッジを保持する同心円画像に対して、2D DWT が分離するエッジの方向を図 2-14(b) に示す。

画像信号は一般的に低周波成分の平坦領域と高周波成分のエッジ信号から構成される。このときエッジ信号はあらゆる方向に発生する可能性があり、これらの方向を適切に分離し解析することが好ましい。しかし、図 2-14(b) のように 2D DWT では、それらを 0° , 90° , $\pm 45^\circ$ の 3 方向にしか分離できず、 $\pm 45^\circ$ 方向については異なる方向のエッジを 1 つのサブバンドに含んでしまっている。これより、2D DWT は画像信号の方向分離特性が乏しい変換といえる。これに対して、第3章で述べる CWT は、6 方向のエッジを異なるサブバンドに分離できる高い方向分離特性をもつ。

2.5 離散ウェーブレット係数の SPIHT 符号化と符号化性能

2D DWT により分割生成されたサブバンドには、サブバンド内およびサブバンド間にある種の冗長性が残っている。第 2.2.3 節で述べた JPEG 2000 はこの冗長性を除去するために EBCOT を用いている。EBCOT はコードブロック内の係数を対象とするため、サブバンド内の係数間に残る冗長性を削減する手法である。EBCOT はサブバンド内で処理が完結するため空間スケーラビリティを提供できるが、階層間の係数に残る冗長性を削減していない。これに対して、本節で述べる SPIHT 符号化はサブバンド内だけでなく、階層間の冗長性も削減するため JPEG 2000 よりも優れた符号化効率を得られる。しかし、空間スケーラ

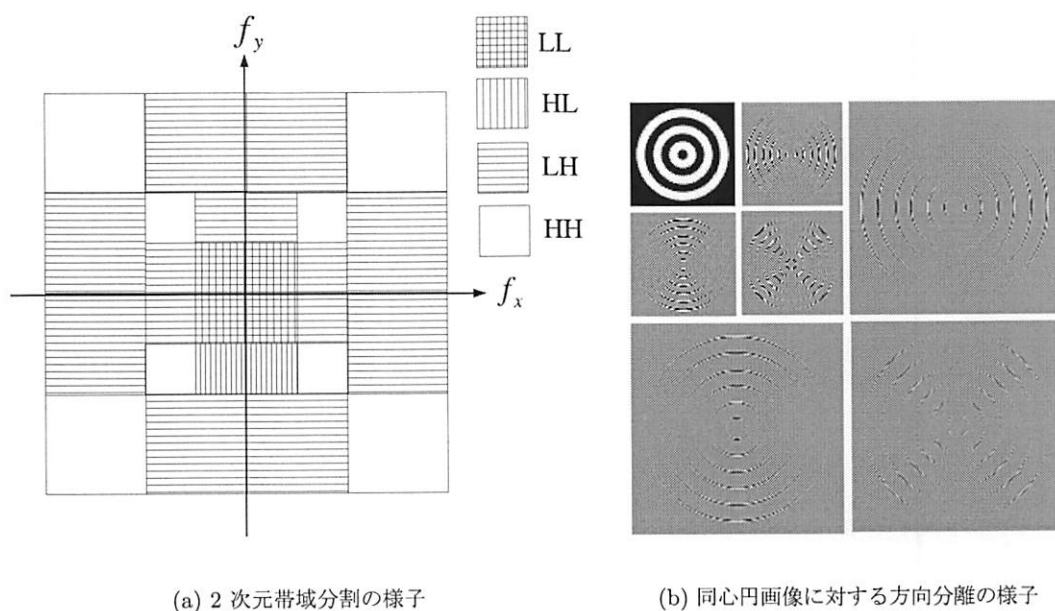


図 2-14: 2次元離散ウェーブレット変換の方向分離特性

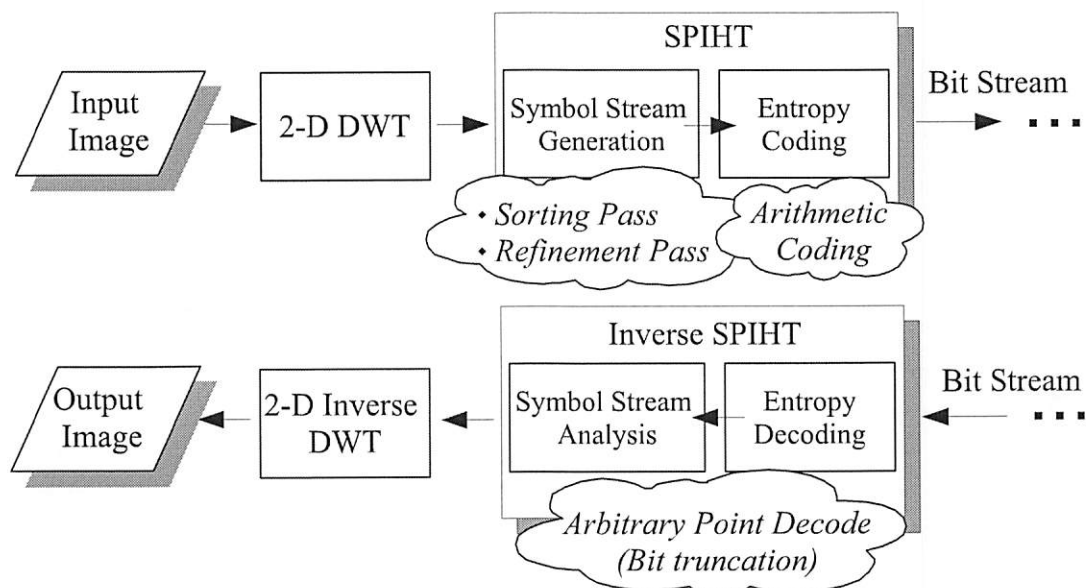
ビリティを提供することはできない。本研究では、DWTによる画像符号化として SPIHT 符号化を用いる。

SPIHT 符号化はゼロツリー構造を用いて冗長性を除去する。この考え方は *Shapiro* により EZW (Embedded Zero-tree coding of Wavelet coefficient)⁽²¹⁾ として最初に提案され、*Said* らにより SPIHT として性能の改善がなされた⁽²²⁾。EZW や SPIHT は任意の受信情報のみから再構成画像を生成できるビットストリームを提供する。このような任意のビットレートで再構成画像を生成できる機能は SNR (Signal to Noise Ratio) スケーラビリティと呼ばれる。

SPIHT の処理フローを図 2-15 に示す。エンコーダではまず入力画像に対して 2D DWT を行う。次に得られたウェーブレット係数に SPIHT を適用する。SPIHT はシンボル生成器とエントロピー符号化器から構成される。シンボル生成器ではゼロツリー構造を用いて効率的にウェーブレット係数の位置情報と係数値情報を符号化する。エントロピー符号化器では算術符号化を用いてシンボルのエントロピー符号化を行う。

デコーダでは、まず受信したビットストリームに対してエントロピー復号化を行う。次にエンコーダで生成されたシンボルをもとにウェーブレット係数の位置情報と係数値を復元する。復元されたウェーブレット係数列に 2D Inverse DWT を行い再構成画像を生成する。シンボルを解析する際にビットランケーションを行うことで、任意のビットレートにおいて再構成画像を復元することができる。

SPIHT では図 2-16 の実線のように異なるレベル間の同空間位置の係数を集合として扱う。集合の各要素は一般的に以下のように呼ばれる。

図 2-15: SPIHT 符号化器と復号化器 (文献⁽²³⁾ 参照)

- Parent : 1 レベル高解像度の同方向・同空間位置の 4 つの係数をまとめる 1 係数
- Child : Parent の定義における 4 つの係数のうちの 1 係数
- Descendants : Parent より高解像度の同方向・同空間位置の係数群
- Root : Descendants を形成する Parent

Parent の絶対値が小さい場合、その Descendants の絶対値も小さいというウェーブレット係数の特性より、上記で定義した集合を利用することで係数の位置情報を効率的に符号化することができる。また SPIHT では図 2-16 の破線矢印のように低解像度レベルから高解像度レベルの順番で処理される。このとき Parent が処理されるまでその Child は処理されないため、Child は Parent に埋め込まれていると考えることができる。一般的に低解像度レベルに属する係数は高解像度レベルに属する係数よりも画像品質に与える影響が大きい。この処理順序は重要度の順に係数列を配置しているといえる。この処理順序に従って Sorting Pass と Refinement Pass と呼ばれる 2 つのパスを交互に適用し、ウェーブレット係数の位置情報と係数値をシンボル化する。

Sorting Pass

Sorting Pass は与えられた閾値を用いて有意係数 (significant coefficients) を判別するパスである。ウェーブレット係数は絶対値の大きい係数ほど画像品質に与える影響が高い。このような係数は有意係数と呼ばれ優先的に符号化する必要がある。まず、全係数の最大値から 2 の階乗に制限される閾値を決定し、これを 1 回目の Sorting Pass の閾値とする。

$$threshold_1 = 2^{\lfloor \log_2(\max|\text{coeff}|) \rfloor} \quad (2.33)$$

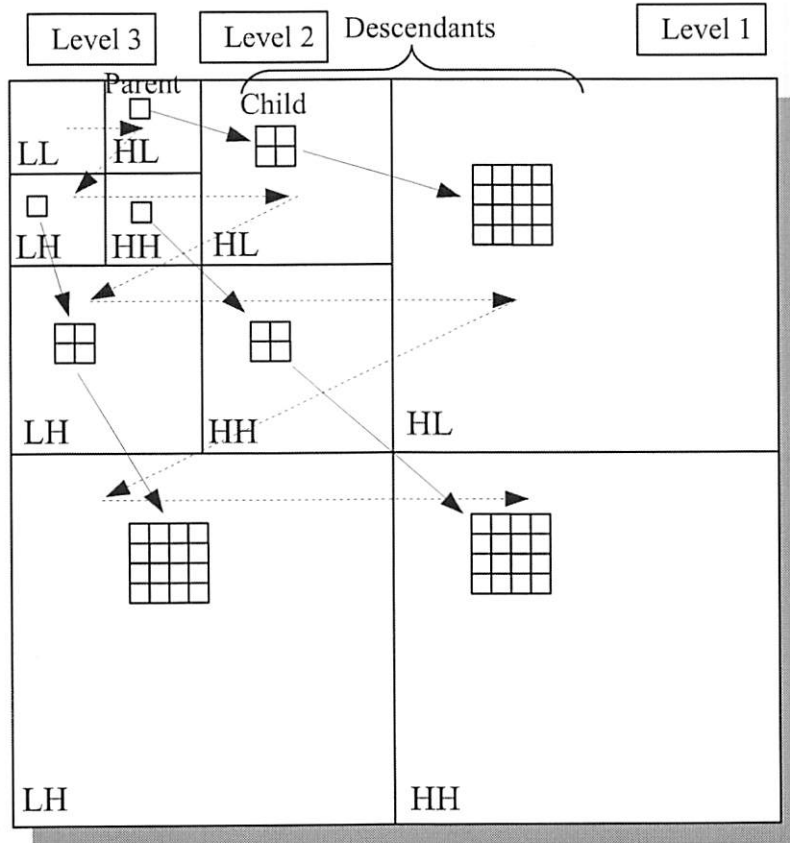


図 2-16: サブバンド上における親子関係

2 回目以降は式 (2.33) の閾値を 2 で割ることによって次第に小さくする. すなわち n 回目の Sorting Pass の閾値は式 (2.34) となる.

$$threshold_n = threshold_1 / 2^{(n-1)} \quad (2.34)$$

Sorting Pass では係数の絶対値 $|c_{ij}|$ と閾値の大きさを比較し, 閾値より大きい係数を有意係数 (1), 小さい係数を非有意係数 (0) とする.

$$S_n(\Gamma) = \begin{cases} 1, & |c_{ij}| \geq threshold_n \\ 0, & \text{otherwise} \end{cases} \quad (2.35)$$

この判定に基づき 4 つのシンボルが生成される⁽²³⁾.

- ZTR (Zero-TRee) シンボル: 対象係数と Descendants のすべてが非有意係数である場合に適用されるシンボルである. このとき, 同方向・同空間のウェーブレット係数の集合を 1 シンボルで効率的に表現できる.
- IZ (Isolated Zero) シンボル: 対象係数が非有意係数であり, Descendants に有意係数が含まれている場合に適用されるシンボルである.

- POS (POSitive) シンボル: 対象係数が有意係数であり, その符号が正である場合に適用されるシンボルである.
- NEG (NEGative) シンボル: 対象係数が有意係数であり, その符号が負である場合に適用されるシンボルである.

POS と NEG シンボルは画像品質に影響を与えるシンボルであり, ZTR と IZ シンボルは効率的に有意係数 (POS, NEG) の位置情報をあらわすために用いられるシンボルである.

Refinement Pass

Refinement Pass は Sorting Pass に続いて処理されるパスである. このパスは $n-1$ 回目の Sorting Pass において有意係数と判定された係数についてビットプレーン展開を用いて係数値の精度を向上させる. その際に次の 2 つのシンボルのいずれかが生成される.

- HIGH シンボル: 対象係数が与えられた閾値より大きい場合に適用されるシンボルである.
- LOW シンボル: 対象係数が与えられた閾値より小さい場合に適用されるシンボルである.

Refinement Pass では図 2-17 のように係数をビットプレーン展開し, MSB (Most Significant Bit) から LSB (Least Significant Bit) の順番に配列する. そして 1 回の Refinement Pass で 1 [bit] ずつ閾値と比較し HIGH(1), LOW(0) いずれかのシンボルを決定する. この処理によりデコーダにおける有意係数値の不確定領域が減少し, 係数値の精度が向上する.

Sorting Pass における 4 つのシンボルと Refinement Pass における 2 つのシンボルの組み合わせがシンボル生成器から出力される. これらのシンボルは算術符号化を用いてエントロピー符号化されビットストリームが形成される.

標準画像 Lena (512×512 サイズ, 8 [bit] グレースケール), Barbara (512×512 サイズ, 8 [bit] グレースケール) および Goldhill (512×512 サイズ, 8 [bit] グレースケール) に対する SPIHT 符号化の性能を図 2-18 に示す. 本研究では, SPIHT 符号化において算術符号化は用いていない. 図 2-18(a)(c)(e) の横軸は 1 画素を符号化するのに必要なビット数である. たとえば 1 [bit/pel] の場合, 1 画素あたり 1 [bit] 必要となるため, 8 [bit] グレースケール画像の場合 1/8 に情報が圧縮されていることを意味する. 縦軸は PSNR (Peak Signal-to-Noise Ratio) であり原画像と劣化画像の差異を客観的に評価する数値である. PSNR は式 (2.36) により定義される.

$$PSNR = 10 \log_{10} \left(\frac{255}{RMSE} \right) \quad (2.36)$$

$$RMSE = \sqrt{\frac{\sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} [f_{n_1 n_2} - \hat{f}_{n_1 n_2}]^2}{N_1 \times N_2}}$$

N_1, N_2 は画像サイズ, f は原画像, $\hat{f}$ は劣化画像である. 図 2-18(b) は 0.2 [bit/pel] における Lena の再構成画像, 図 2-18(d) は 0.5 [bit/pel] における Barbara の再構成画像, 図 2-18(f) は 0.3 [bit/pel] における Goldhill の再構成画像である. 各再構成画像において, 対

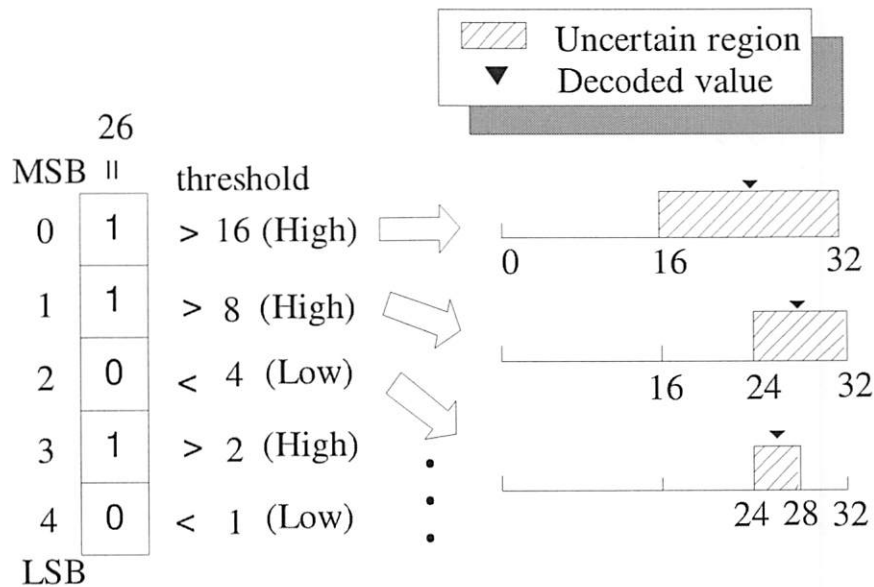


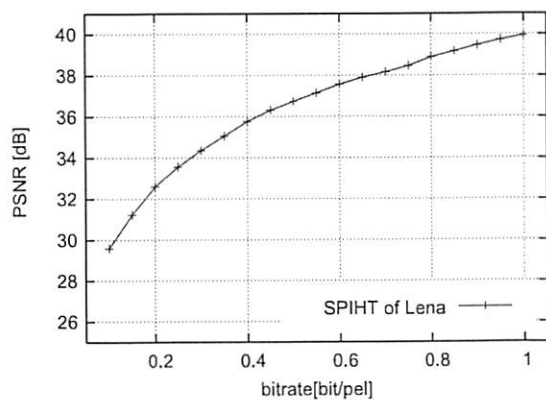
図 2-17: ビットプレーン展開と量子化

角方向付近のエッジ上に階段状のアーティファクトが生じていることがわかる。このアーティファクトはチェッカーボードアーティファクトと呼ばれ、第 2.4.2 節で述べた DWT の方向分離特性が影響している。2D DWT は、HH サブバンドが $\pm 45^\circ$ 方向の 2 方向の周波数帯域を含んでいる。そのため、この帯域の係数値を量子化により変更すると、このようなチェッカーボード状のアーティファクトが生じる。したがって、このアーティファクトを回避するためには、各サブバンドが単一方向のエッジにのみ反応する方向分離特性を持つ必要がある。第 3 章で述べる DTCWT はこの特性を持つ変換であるため、チェッカーボードアーティファクトを回避できる。

2.6 むすび

本章では静止画像の変換符号化の現状について述べた。まず、変換符号化を用いた国際標準画像符号化 JPEG と JPEG 2000 の特徴と技術について述べた。次に、JPEG 2000 の変換手法である DWT について述べた。そして、DWT の問題点を示した。最後に、DWT の SPIHT 符号化について述べ、その符号化性能を示した。

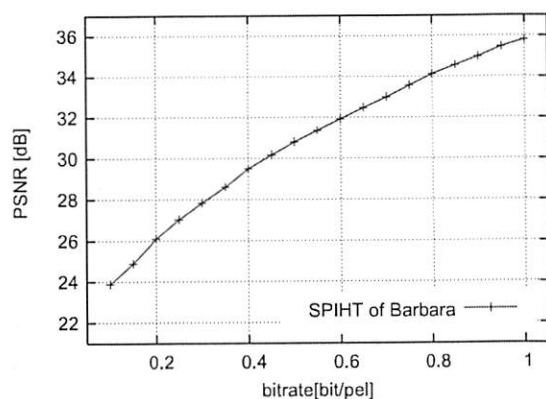
第 3 章では、DWT と同じウェーブレット変換であるが、その特性が大きく異なる DTCWT について述べ、DTCWT を用いた画像符号化について言及する。



(a) Lena のレート歪み特性



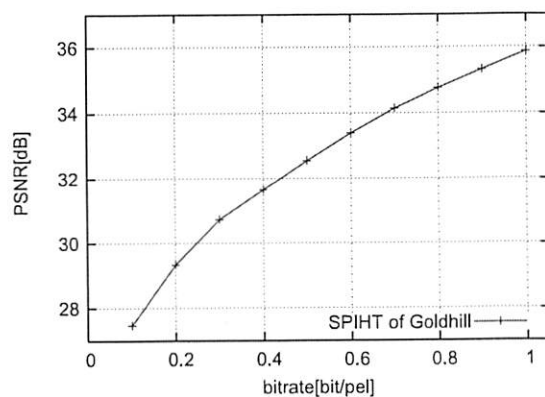
(b) 0.2 (bit/pel) : 32.61 (dB)



(c) Barbara のレート歪み特性



(d) 0.5 (bit/pel) : 30.81 (dB)



(e) Goldhill のレート歪み特性



(f) 0.3 (bit/pel) : 30.73 (dB)

図 2-18: 離散ウェーブレット係数に対する SPIHT 符号化の符号化性能

第3章

並列木複素ウェーブレット変換による新しい静止画像符号化

3.1 まえがき

近年、離散系において DTCWT と呼ばれるウェーブレット変換が提案されている。DTCWT は第 2.4 節で述べた DWT の問題点を解決する。すなわち、DTCWT はシフト不変性が成立し、2 次元変換は $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ の 6 方向のエッジを異なるサブバンドに分離する高い方向分離特性をもつ。また、DTCWT は変換係数が複素数となる複素変換であり、そのデータ数は変換前のデータ数より増加する。これは実数変換であり変換前後のデータ数が不変である DWT と対称的である。DTCWT のような変換後のデータ数が増加する冗長変換では、疎表現と呼ばれる少数の基底により高い精度で信号を近似する係数列を求めることができる。本研究は、DTCWT の疎表現に基づく静止画像符号化を実現するという点で、国際標準符号化と大きく異なっている。

本章では、DTCWT による新しい静止画像符号化について述べる。まず、2 種類の複素ウェーブレット変換 (CWT : Complex Wavelet Transform) があることを述べる。次に、ヒルベルト変換について説明し、FT と複素ウェーブレットの類似性を述べる。そして、DTCWT のフィルタバンクによる実装方法やフィルタの設計方法、さらに 2 次元への拡張方法について述べる。次に、DTCWT の特性であるシフト不変性の成立と 2 次元変換における高性能な方向分離特性を示す。そして、DTCWT 係数に残る冗長性について示し、オリジナルの DTCWT 係数列を用いた SPIHT 符号化では高い符号化効率が得られないことを確認し、疎表現の必要性を述べる。最後に、冗長変換における疎表現の理論と疎表現を求める近似解法について述べる。本章で述べる内容に基づき第 4 章において、DTCWT の疎表現を用いた静止画像符号化の検討を行う。

3.2 複素ウェーブレット変換の種類

CWT は 2 種類に大別することができる。1 つは冗長複素ウェーブレット変換 (RCWT : Redundant CWT) であり、もう 1 つは非冗長複素ウェーブレット変換 (NRCWT : Non-Redundant CWT) である。RCWT は N サンプルの入力信号に対して、変換後のサンプル

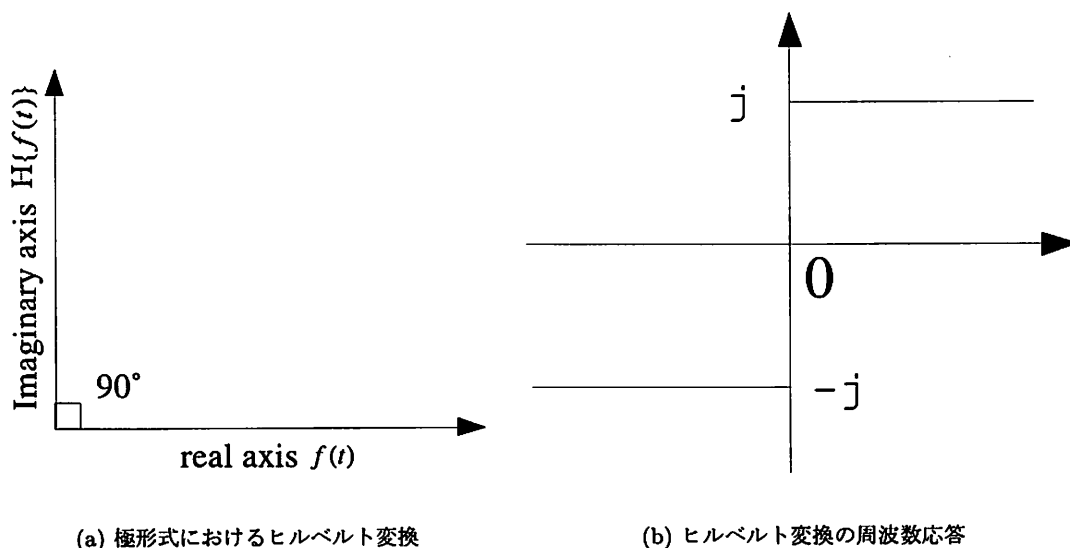


図 3-1: ヒルベルト変換

数が $M (M > N)$ となる CWT である。すなわち、変換後のデータ数が増加する変換である。これに対して NRCWT は N サンプルの入力信号に対して、変換後のサンプル数が N となる CWT である。すなわち、変換前後でデータ数が不変の変換である。

RCWT は 1998 年に *Kingsbury* により提案された。構造は 2 つの DWT のフィルタバンクを並列に配置し、一方を実部のフィルタバンク、他方を虚部のフィルタバンクとする並列木構造である。そのため並列木複素ウェーブレット変換 (DTCWT) と呼ばれる。DTCWT はシフト不変性が成立し、2 次元変換では $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ の 6 方向のエッジを異なるサブバンドに分離する特性がある⁽¹⁹⁾⁽²⁶⁾⁽²⁷⁾。この変換ではヒルベルト変換対をなす 2 つのウェーブレットを用いる⁽²⁸⁾。ヒルベルト変換については第 3.3 節で述べる。また、第 3.5.2 節において DTCWT におけるウェーブレットの設計方法を 2 種類述べる。これ以外の設計方法として、2 種類の Spline ウェーブレットを実部と虚部に用いる RI-Spline ウェーブレットなどがある⁽²⁹⁾。本研究では DTCWT を用いている。

NRCWT は 2002 年に *Fernandes* らによりマッピングベース複素ウェーブレット変換 (MBCWT : Mapping-Based CWT) として提案された⁽³⁰⁾⁽³¹⁾。この変換はデジタルフィルタを通して実信号を解析信号に変換し、変換された解析信号に対して DWT を行う手法である。解析信号への変換と DWT を分離して処理できるため、解析信号への変換の際に、冗長度を制御することができる。冗長度とシフト不変性はトレードオフの関係があるため、MBCWT ではシフト不変性が不成立な 6 方向分離特性を持つ NRCWT を実現することができる。また、解析信号への変換と DWT が独立に行われる特性を利用して、DWT を倍密度 DWT (Double-Density DWT) に置き換えた複素倍密度 DWT (Complex Double-Density DWT) も提案されている⁽³⁰⁾。

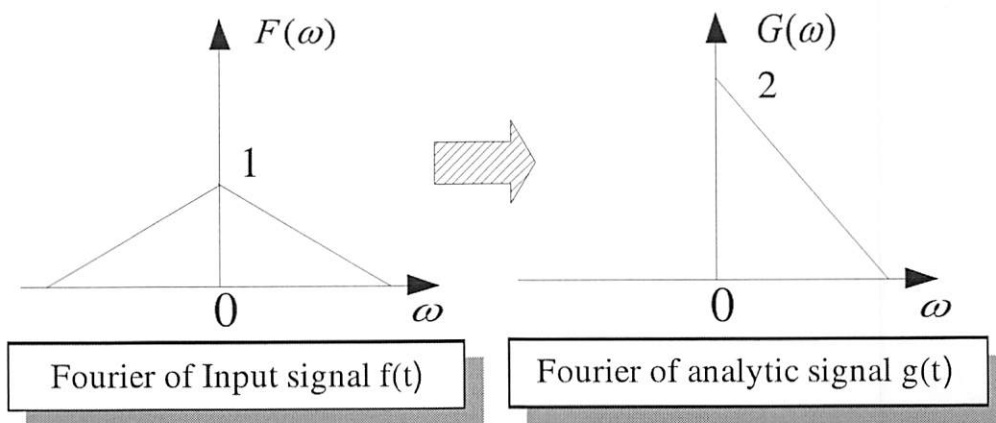


図 3-2: 解析信号の周波数表現

3.3 ヒルベルト変換と解析信号

ヒルベルト変換 (hilbert transform) は信号の全周波数成分の位相を 90° 遅らせる変換である⁽³²⁾⁽³³⁾. 信号 $f(t)$ に対するヒルベルト変換の時間表現は式 (3.1) により与えられる.

$$\mathcal{H}\{f(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(\tau)}{t - \tau} d\tau = f(t) * \frac{1}{\pi t} \quad (3.1)$$

ここで $*$ は畳み込み計算であり式 (3.2) で定義される.

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(u)y(t - u)du \quad (3.2)$$

$f(t)$ の FT を $F(\omega)$ とするとヒルベルト変換の周波数表現は式 (3.3) となる.

$$F[\mathcal{H}\{f(t)\}] = \begin{cases} -j F(\omega), & \omega > 0 \\ 0, & \omega = 0 \\ j F(\omega), & \omega < 0 \end{cases} \quad (3.3)$$

ここで $j = \sqrt{-1}$ である. 極形式におけるヒルベルト変換の関係を図 3-1(a), また, ヒルベルト変換の周波数応答を図 3-1(b) に示す. 図 3-1(b) よりヒルベルト変換は正周波数成分の位相を 90° 遅らせ, 負周波数成分の位相を 90° 進める機能を持つフィルタである.

信号 $f(t)$ とそのヒルベルト変換 $\mathcal{H}\{f(t)\}$ を用いて解析信号 (analytic signal) $g(t)$ を式 (3.4) により定義する.

$$g(t) = f(t) + j\mathcal{H}\{f(t)\} \quad (3.4)$$

このように定義された解析信号 $g(t)$ は $f(t)$ を実部, $\mathcal{H}\{f(t)\}$ を虚部とする複素関数とみなすことができる. つまり, 実関数 $f(t)$ からそのヒルベルト変換を経由して複素関数 $g(t)$ を作り出せる. 解析信号は, 実数信号における負成分周波数が正周波数成分へ折り返されるため, 正周波数成分のみをもつ. これは複素数信号の FT が片側周波数成分のみをもち, 実数

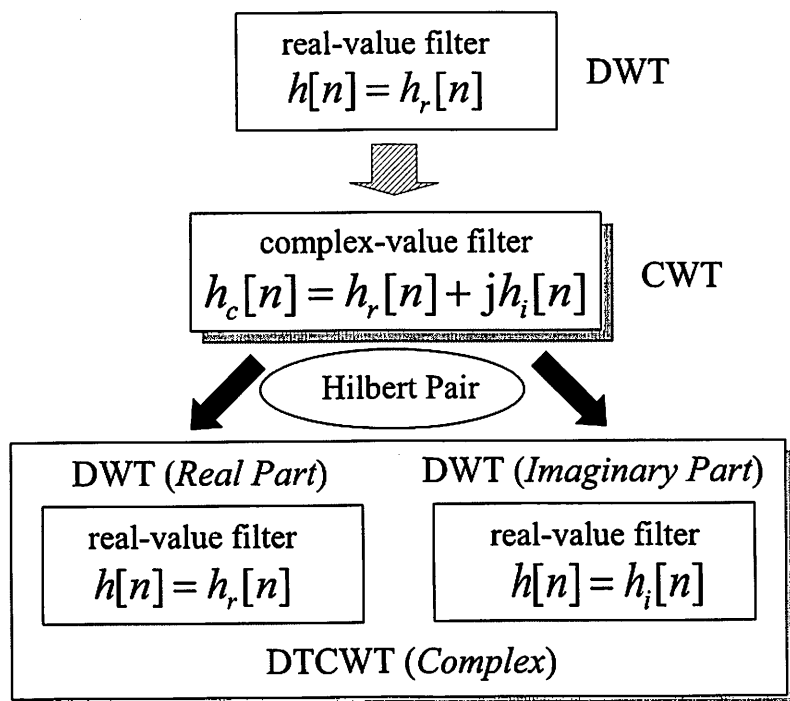


図 3-3: 複素ウェーブレット変換の概念

信号の FT が対称な正負周波数成分をもつことから理解される。この様子を図 3-2 に示す。解析信号を用いることで、片側周波数成分のみに帯域制限されたフィルタを設計することができる。CWT は解析信号のこの特性を利用して、シフト不変性欠如の原因となるダウンサンプリングに伴うエイリアシングの発生を防いでいる。

DWT のフィルタバンクに解析信号を用いると CWT が実現できる。この様子を図 3-3 に示す。まず、DWT のフィルタバンクで用いる実数フィルタを複素数フィルタに置き換える。次に、複素数フィルタを 2 つの実数フィルタへ分解する。このとき分解された 2 つの実数フィルタはヒルベルト変換対を形成しており、これらのフィルタペアから形成される複素数フィルタは解析フィルタとなっている。各実数フィルタでは独立に DWT を実行し、得られたウェーブレット係数を実部係数、虚部係数とすることで、全体として複素変換を実現する。

3.4 複素ウェーブレットとフーリエ変換

第 2.4 節においてシフト不変性の欠如と 2 次元信号における低性能な方向分離特性という DWT の問題点を述べた。これらの問題は FT では起こらない。すなわち、FT の絶対値はシフト不変性が成立し、2 次元変換は高い方向分離特性を持つ。FT のシフト不変性については式 (3-5) から理解できる。

$$f(t - t_0) \Rightarrow F(\omega)e^{-j\omega t_0} \quad (3-5)$$

式(3.5)は入力信号 $f(t)$ の時間シフト t_0 に対して, FT は位相の回転のみで絶対値は不変であることを示している. FT の高性能な方向分離特性については第3.6.2節において CWT の方向分離特性とあわせて述べる.

DWT と FT の違いは積分計算に用いる関数である⁽³³⁾. DWT では局所的に振動する実数関数を用いるが, FT では式(3.6)の複素正弦波を用いる.

$$e^{j\omega t} = \cos(\omega t) + j \sin(\omega t) \quad (3.6)$$

$\cos$ 関数と $\sin$ 関数は第3.3節で述べたヒルベルト変換対となっており, これらから構成される $e^{j\omega t}$ は解析信号である. つまり FT は解析信号を用いた変換である. FT と同様に解析信号となるウェーブレットを用いてウェーブレット変換を行う手法が CWT である. ヒルベルト変換対を満たす2つのウェーブレット ψ^r , ψ^i をそれぞれ実部, 虚部として, 複素ウェーブレット ψ^c を式(3.7)で定義する. このとき ψ^c は解析信号である.

$$\psi^c(t) = \psi^r(t) + j\psi^i(t) \quad (3.7)$$

複素スケーリング関数 φ^c についても同様に定義される. このように定義した複素ウェーブレットを $\psi_{j,k}^c = 2^{j/2}\psi^c(2^j t - k)$ のように離散化すると, 式(2.18)の DWT と同様に関数の再構成表現を定義できる.

$$f(t) = \sum_{k=-\infty}^{\infty} C_c(J_0, k) \varphi_{J_0, k}^c(t) + \sum_{j=J_0}^{J_1} \sum_{k=-\infty}^{\infty} D_c(j, k) \psi_{j, k}^c(t) \quad (3.8)$$

ここで, 複素ウェーブレット係数 $D_c(j, k)$ は実部と虚部のウェーブレット係数 $D_r(j, k)$, $D_i(j, k)$ と式(3.9)の関係となっている.

$$D_c(j, k) = D_r(j, k) + jD_i(j, k) \quad (3.9)$$

$D_c(j, k)$ は複素数であるため, 絶対値と位相を式(3.10)と式(3.11)により定義できる. FT と同様に, 複素ウェーブレット係数の絶対値は特異点の存在を示し, 位相はその位置を示す. 複素スケーリング係数 $C_c(J_0, k)$ についても同様である.

$$|D_c(j, k)| = \sqrt{|D_r(j, k)|^2 + |D_i(j, k)|^2} \quad (3.10)$$

$$\angle D_c(j, k) = \arctan\left(\frac{D_i(j, k)}{D_r(j, k)}\right) \quad (3.11)$$

3.5 並列木複素ウェーブレット変換

3.5.1 並列木フィルタバンクによる実装

RCWT である DTCWT は DWT と同様にフィルタバンクにより実装される. 1 次元 DTCWT (1D DTCWT) の分割・合成フィルタバンクの構造は図3-4, 図3-5となる⁽¹⁹⁾⁽³³⁾. 分割フィルタバンクは2つの DWT のフィルタバンクを並列に配置した構造である. 各フィ

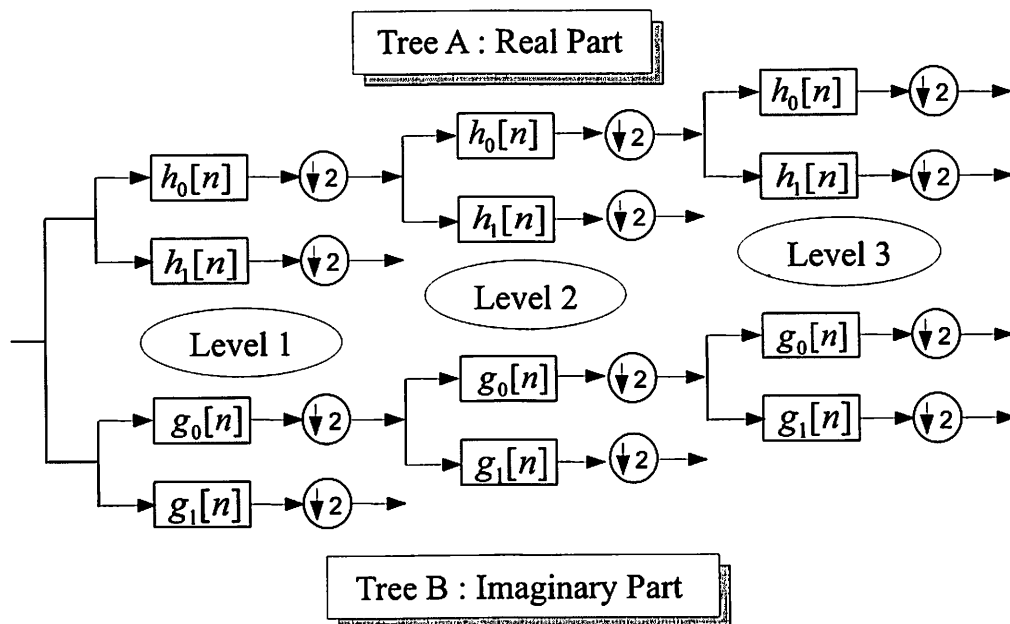


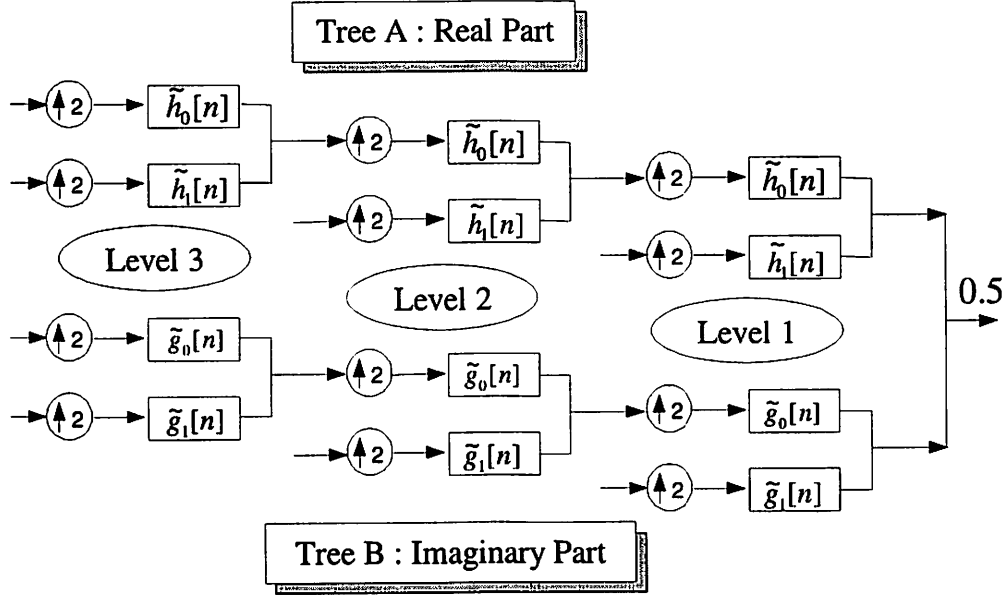
図 3-4: 1次元並列木複素ウェーブレット変換の分割フィルタバンク (文献⁽³³⁾ 参照)

ルタバンクには同じ信号が入力され、独立に DWT の分割処理が行われる。また、各フィルタバンクにおいて完全再構成が成り立つ。2つのフィルタバンクからの出力結果は実部、虚部として扱い、並列木フィルタバンクは全体として複素係数を生成するフィルタバンクとなる。DTCWT は 2つの DWT の最大間引きフィルタバンクを用いるため 1次元において 2:1 の冗長性が発生する。ここでは複素数の実部の役割をするフィルタバンクを TreeA、虚部の役割をするフィルタバンクを TreeB とする。1D DTCWT の複素フィルタは $h_x[n] + jg_x[n]$ であり、 $h_x[n]$ は実部フィルタの組 $\{h_0[n], h_1[n]\}$ 、 $g_x[n]$ は虚部フィルタの組 $\{g_0[n], g_1[n]\}$ である。ここで、 $h_0[n]$ 、 $g_0[n]$ はローパスフィルタ、 $h_1[n]$ 、 $g_1[n]$ はハイパスフィルタである。

合成フィルタバンクでは各フィルタバンクで独立に合成処理を行った後、各々の合成信号を平均することで最終的な合成信号を得る。合成フィルタバンクで用いるフィルタの組 $\{\tilde{h}_0[n], \tilde{h}_1[n]\}$ と $\{\tilde{g}_0[n], \tilde{g}_1[n]\}$ は分割フィルタバンクのフィルタの組 $\{h_0[n], h_1[n]\}$ 、 $\{g_0[n], g_1[n]\}$ とそれぞれ直交または双直交となっている。

3.5.2 ヒルベルト変換対のフィルタ設計

コンパクトサポートを持つウェーブレットを用いた複素ウェーブレットは厳密な解析信号とはなれず、近似的な解析信号しか与えない⁽³³⁾。そのため複素ウェーブレットをできるだけ解析信号に近づける必要がある。言い換えれば、実部と虚部のウェーブレット間に近似精度の良いヒルベルト変換を成立させる必要がある。ウェーブレットの設計問題はフィルタの設計問題に置き換えることができる。ウェーブレットをヒルベルト変換対とするためには、図 3-4 の 2つのフィルタバンク間で、ある関係性が成立するフィルタを設計する必要がある。

図 3-5: 1次元並列木複素ウェーブレット変換の合成フィルタバンク (文献⁽³³⁾ 参照)

式 (2.16) と式 (2.17) のトゥースケール関係よりウェーブレットはローパスフィルタ $h_0[n]$ に依存している。Selesnick はローパスフィルタ間に $1/2$ サンプルの遅延関係が成立するときウェーブレット間にヒルベルト変換対が成り立つことを示した⁽²⁸⁾。

$$g_0[n] \approx h_0[n - 0.5] \Rightarrow \psi_g(t) \approx \mathcal{H}\{\psi_h(t)\} \quad (3.12)$$

これは周波数表現では式 (3.13) となる。

$$G_0(e^{j\omega}) = e^{-j0.5\omega} H_0(e^{j\omega}) \Rightarrow \psi_g(t) = \mathcal{H}\{\psi_h(t)\} \quad (3.13)$$

式 (3.13) の逆条件については文献⁽³⁴⁾⁽³⁵⁾ により証明され、式 (3.13) は必要十分条件となっている。また、双直交の場合についても必要十分条件が証明されている⁽³⁶⁾。式 (3.12) の $1/2$ サンプルの遅延関係は、各レベルにおいてローパス信号を $2:1$ にオーバーサンプリングしていることと等価である。よって、DTCWT ではオーバーサンプリングにより、DWT のダウンサンプリングに伴い発生するエイリアシングが低減され、シフト不変性が近似的に成立しているとも言える⁽²⁶⁾⁽²⁷⁾⁽³⁷⁾。ここで、式 (3.13) のローパスフィルタ間の周波数条件を絶対値と位相に分けて記述すると式 (3.14) と式 (3.15) になる。

$$|G_0(e^{j\omega})| = |H_0(e^{j\omega})| \quad (3.14)$$

$$\angle G_0(e^{j\omega}) = \angle H_0(e^{j\omega}) - 0.5\omega \quad (3.15)$$

よって、ヒルベルト変換対の条件である式 (3.14)、式 (3.15) と完全再構成を同時に満たす FIR ローパスフィルタ $h_0[n]$ 、 $g_0[n]$ を設計する必要がある。これを実現するための 2 種類の設計手法を以下に述べる⁽³³⁾。

LINEAR-PHASE BIORTHOGONAL SOLUTION

この設計方法では、 $h_0[n]$ として奇数長偶対称フィルタ、 $g_0[n]$ として偶数長偶対称フィルタを用いる。

$$h_0[n] = h_0[N-1-n] \quad (3-16)$$

$$g_0[n] = g_0[N-n] \quad (3-17)$$

ここで N はフィルタ長であり奇数である。このとき両フィルタは対称型であるため線形位相となるが、線形位相 FIR フィルタは直交フィルタとならないため、2つのフィルタは双直交フィルタとして設計される。 $h_0[n]$ の位相は $\angle H_0(e^{j\omega}) = -0.5(N-1)\omega$ 、 $g_0[n]$ の位相は $\angle G_0(e^{j\omega}) = -0.5N\omega$ となる。したがって、この設計方法では2つのフィルタの位相は正確に $1/2$ ずれる。しかし、フィルタのタップ長が異なるため周波数応答の絶対値を完全に一致させることはできない。以上より、この設計方法では周波数上で式 (3-18) の関係となる。

$$\begin{aligned} |G_0(e^{j\omega})| &\neq |H_0(e^{j\omega})|, \\ \angle G_0(e^{j\omega}) &= \angle H_0(e^{j\omega}) - 0.5\omega \end{aligned} \quad (3-18)$$

よって、 $h_0[n]$ と $g_0[n]$ の周波数応答の絶対値が近似的に式 (3-14) を満たすように設計しなければならない。

Q-SHIFT SOLUTION

この設計方法では、ローパスフィルタ係数間に時間反転の関係を持たせる。

$$g_0[n] = h_0[N-1-n] \quad (3-19)$$

ここで N は偶数であり $h_0[n]$ のフィルタ長である。この設計方法では係数間の時間反転の関係より周波数応答の絶対値は一致するが、位相の $1/2$ ずれが成立しない。

$$\begin{aligned} |G_0(e^{j\omega})| &= |H_0(e^{j\omega})|, \\ \angle G_0(e^{j\omega}) &\neq \angle H_0(e^{j\omega}) - 0.5\omega \end{aligned} \quad (3-20)$$

よって、位相間に式 (3-14) の関係を近似的に成立させる必要がある。ここで式 (3-19) の FT は式 (3-21) となる。

$$G_0(e^{j\omega}) = \overline{H_0(e^{j\omega})} e^{-j(N-1)\omega} \quad (3-21)$$

- は共役複素数をあらわしている。式 (3-21) を位相関係で表現すると式 (3-22) となる。

$$\angle G_0(e^{j\omega}) = -\angle H_0(e^{j\omega}) - (N-1)\omega \quad (3-22)$$

ここで、2つのフィルタが式 (3-15) の位相関係を近似的に満たしているとする、これを式 (3-22) に代入し式 (3-23) となる。

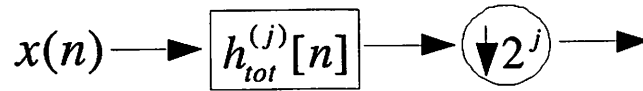
$$\angle H_0(e^{j\omega}) - 0.5\omega \approx -\angle H_0(e^{j\omega}) - (N-1)\omega,$$

$$\angle H_0(e^{j\omega}) \approx -0.5(N-1)\omega + 0.25\omega \quad (3.23)$$

式 (3.23) より $h_0[n]$ は近似的な線形位相フィルタであり, $n = 0.5(N-1) - 0.25$ を中心とした対称型となっている. これは通常の対称点から $1/4$ だけシフトした形式となっている. このため, この種のフィルタは q-shift フィルタと呼ばれる. また, q-shift フィルタは近似的に線形位相を満たすため厳密な対称性を要求されない. そのため, 直交 FIR フィルタとして設計することが可能である.

3.5.3 並列木フィルタバンクの制約

DTCWT は図 3-4 のフィルタバンクを用いて実現されるが, このときレベル 2 以降で精度の良い解析信号を生成するため, レベル 1 を犠牲にしなければならない. つまり, DTCWT ではレベル 1 において解析信号を生成できないという制約が存在する. 式 (3.12) の $1/2$ サンプルの遅延関係を満たすとき, スケーリング関数の間においても $\varphi_g(t) \approx \varphi_h(t-0.5)$ の $1/2$ 遅延関係が成立する⁽³³⁾. 離散系においてスケーリング関数は整数値でシフトするので, この関係は 2 つのスケーリング関数が相互に補間する位置関係にあることを意味している. 図 3-4 における TreeA のレベル j までのローパスフィルタのフィルタリング処理をまとめてあらわす.



ここで, $h_{tot}^{(j)}[n]$ の周波数応答は次式である.

$$H_{tot}^{(j)}(z) = H_0^{(1)}(z)H_0^{(2)}(z^2)\dots H_0^{(j)}(z^{2^{j-1}}) \quad (3.24)$$

$H_0^{(j)}(z^{2^{j-1}})$ は, レベル j におけるローパスフィルタの周波数応答である. TreeB のレベル j までのローパスフィルタのフィルタリング処理に関しても同様に $g_{tot}^{(j)}[n]$ と $G_{tot}^{(j)}(z)$ を定義する. 各レベルにおいてスケーリング関数が相互補間の位置関係となるためには, レベル j において 2^j でシフトする $g_{tot}^{(j)}[n]$ が, 同じく 2^j でシフトする $h_{tot}^{(j)}[n]$ の中間に位置する必要がある. レベル 1, レベル 2, およびレベル 3 における, この関係を式 (3.25) に示す.

$$g_{tot}^{(1)}[n] \approx h_{tot}^{(1)}[n-1], \quad g_{tot}^{(3)}[n] \approx h_{tot}^{(3)}[n-2], \quad g_{tot}^{(2)}[n] \approx h_{tot}^{(2)}[n-4] \quad (3.25)$$

レベル 1 のとき, $h_{tot}^{(1)}[n] = h_0^{(1)}[n]$, $g_{tot}^{(1)}[n] = g_0^{(1)}[n]$ より,

$$g_0^{(1)}[n] \approx h_0^{(1)}[n-1] \quad (3.26)$$

となる. すなわち, レベル 1 については 2 つのローパスフィルタが $1/2$ サンプルの遅延関係を満たさない. このためレベル 1 のサブバンドは解析信号を与えることができない. また, 式 (3.26) よりレベル 1 では両 Tree で同じフィルタを用いて, 1 サンプルのオフセットを与えたフィルタリング処理をすればよい.

表 3-1: 並列木複素ウェーブレット変換のフィルタ係数

Level 1		Remaining levels			
Tree A, Tree B		Tree A		Tree B	
$h_0[n]$	$h_1[n]$	$h_0[n]$	$h_1[n]$	$g_0[n]$	$g_1[n]$
-0.0018	-0.0001	0.0024	0.0023	-0.0023	0.0024
0.0000	0.0000	0.0013	0.0012	0.0012	-0.0013
0.0223	0.0013	-0.0026	0.0118	-0.0118	-0.0026
-0.0469	-0.0019	-0.0066	0.0013	0.0013	0.0066
-0.0482	-0.0072	0.0315	-0.0444	0.0444	0.0315
0.2969	0.0239	0.0182	-0.0533	-0.0533	-0.0182
0.5555	0.0556	-0.1202	0.1133	-0.1133	-0.1202
0.2969	-0.0517	0.0246	0.2809	0.2809	-0.0246
-0.0482	-0.2998	0.5658	-0.7528	0.7528	0.5658
-0.0469	0.5594	0.7528	0.5658	0.5658	-0.7528
0.0223	-0.2998	0.2809	-0.0246	0.0246	0.2809
0.0000	-0.0517	-0.1133	-0.1202	-0.1202	0.1133
-0.0018	0.0556	-0.0533	-0.0182	0.0182	-0.0533
	0.0239	0.0444	0.0315	0.0315	-0.0444
	-0.0072	0.0013	0.0066	-0.0066	0.0013
	-0.0019	-0.0118	-0.0026	-0.0026	0.0118
	0.0013	0.0012	-0.0013	0.0013	0.0012
	0.0000	-0.0023	0.0024	0.0024	0.0023
	-0.0001				

レベル 2 以降については $1/2$ サンプルの遅延関係が成立する。レベル 2 に関して、これを示す。レベル 2 における 2 つのローパスフィルタ間の補間関係を FT すると式 (3-27) となる。

$$G_{tot}^{(2)}(e^{j\omega}) \approx e^{-j2\omega} H_{tot}^{(2)}(e^{j\omega}) \quad (3-27)$$

式 (3-24) の関係を式 (3-27) に代入すると次式となる。

$$G_0^{(1)}(e^{j\omega}) G_0^{(2)}(e^{j2\omega}) \approx e^{-j2\omega} H_0^{(1)}(e^{j\omega}) H_0^{(2)}(e^{j2\omega}) \quad (3-28)$$

式 (3-26) より $G_0^{(1)}(e^{j\omega}) \approx e^{-j\omega} H_0^{(1)}(e^{j\omega})$ であるから、式 (3-28) は次式のように変形できる。

$$G_0^{(2)}(e^{j2\omega}) \approx e^{-j\omega} H_0^{(2)}(e^{j2\omega})$$

$$G_0^{(2)}(e^{j\omega}) \approx e^{-j0.5\omega} H_0^{(2)}(e^{j\omega}) \quad (3-29)$$

式 (3-29) は時間表現では $g_0^{(2)}[n] = h_0^{(2)}[n - 0.5]$ となり、 $1/2$ サンプルの遅延条件を満たしている。レベル 3 以降についても同様に示すことができる。

以上より、レベル 1 は同じフィルタを用いて実装され、生成されるサブバンドは解析信号とならない。レベル 2 以降では、 $1/2$ サンプルの遅延関係を満たす 2 種類のフィルタを用いて実装され、生成されるサブバンドは解析信号となる。なお、本研究では DTCWT で用いるフィルタとして、レベル 1 において 13-19 タップ奇対称型フィルタを用いる。レベル 2 以降において“Q-SHIFT SOLUTION”により作られる 18 タップ q-shift フィルタとそれを時間反転したフィルタの 2 種類を用いる。これらのフィルタ係数を表 3-1 に示す。

3.5.4 2次元並列木複素ウェーブレット変換

DWT と同様に 2 次元 DTCWT (2D DTCWT) は 1D DTCWT を拡張したものである。2D DTCWT の分割フィルタバンクの構成を図 3-6 に示す。2D DWT と同様に横方向・縦方向の順で処理が進められる。まず、レベル 1 では入力画像に対して図 3-4 と同様の横方向フィルタリングを行い TreeA と TreeB を生成する。次に、それぞれの Tree に対して縦方向フィルタリングを実行し TreeA*TreeA, TreeB*TreeA, TreeA*TreeB, TreeB*TreeB の 4 つの Tree を生成する。レベル 2 以降ではレベル 1 において生成された 4 つの Tree の LL 成分に対して、2D DWT と同様のフィルタリング処理を行う。この結果、各 Tree から 2D-DWT と同様の LL, LH, HL, HH サブバンドができるため、合計で 4 つの LL, HL, LH, HH サブバンドができる。よって、2D DTCWT は 2D DWT の 4 倍のデータ数が発生し、4:1 の冗長性を持つ変換となる。最後に複素数を形成するために、対応するサブバンド間で加算処理と減算処理を行い、6 方向のエッジを分離するサブバンド $D_\theta^c = D_\theta^r + jD_\theta^i : \theta = \{\pm 75^\circ, \pm 45^\circ, \pm 15^\circ\}$ ができる。どのような加算処理と減算処理が必要かについては第 3.6.2 節で述べる。2D DTCWT の合成処理は分割処理の逆処理として 1D DTCWT の合成処理の拡張として実装される。図 3-7 に自然画像 Lena (512×512 , 8[bit] グレースケール) と Barbara (512×512 , 8[bit] グレースケール) に対する 2D DTCWT を示す。図 3-7(a)(c) は DTCWT 係数の絶対値、図 3-7(b)(d) は DTCWT 係数の位相情報から再構成された画像である。2D DTCWT 係数の絶対値は画像中に含まれるエッジ等の特異点の大きさをあらわす。位相は特異点の位置を示すため、位相再構成画像は画像中の幾何構造をあらわしている。

3.6 並列木複素ウェーブレット変換の特性

3.6.1 近似的シフト不変性の成立

DTCWT は第 2.4 節で述べた DWT の問題点であるシフト不変性の欠如を改善することができる。この様子を図 3-8 に示す。図 3-8(a) は各レベルにおける DTCWT 係数のインパルス応答のエネルギー、図 2-13(b) はレベル 3 における DTCWT 係数のインパルス応答の波形である。インパルス信号のシフトに対して、各レベルのエネルギーはほとんど変化せず、インパルス応答の波形の形状もほぼ一定となっている。これより DTCWT では近似的にシフト不変性が成立していることがわかる。

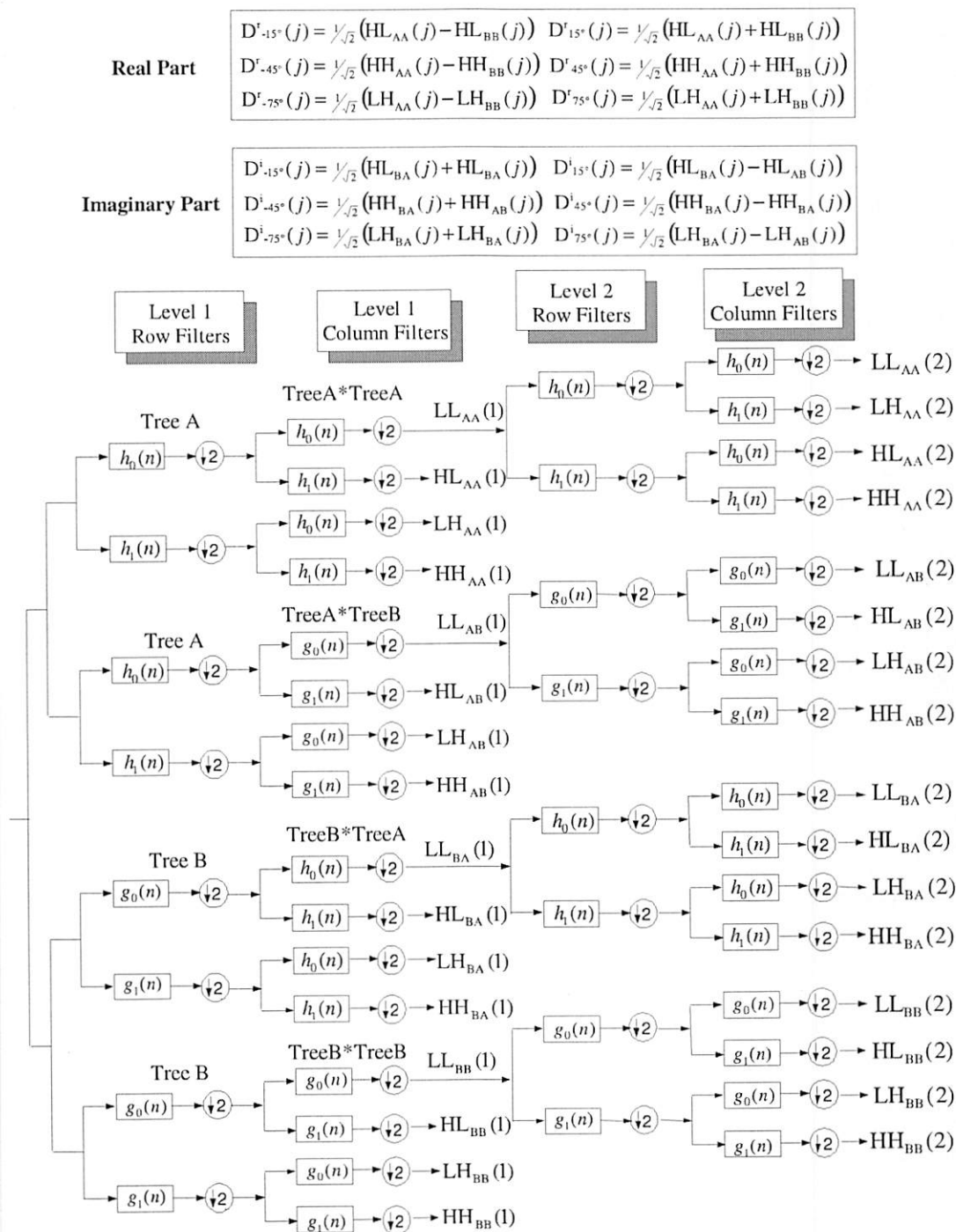
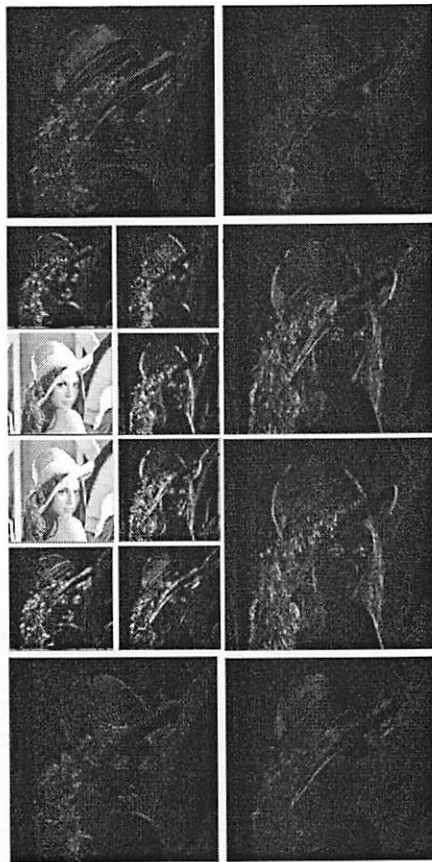
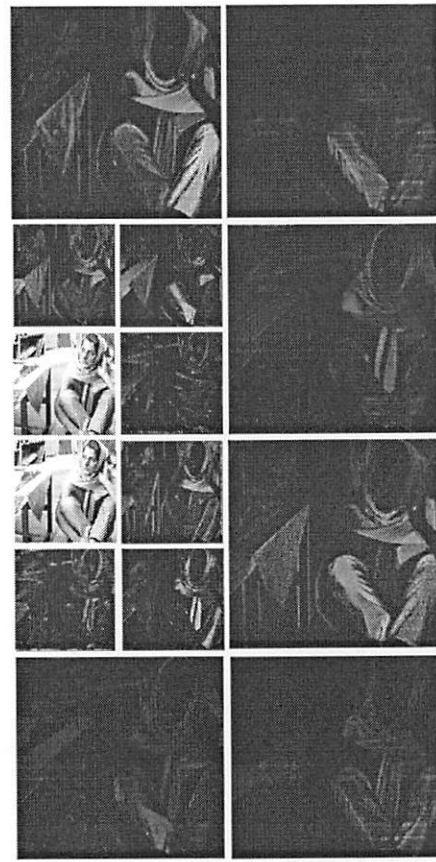


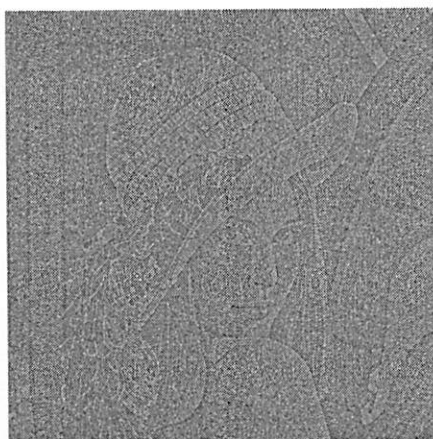
図 3-6: 2次元並列木複素ウェーブレット変換の分割フィルタバンク



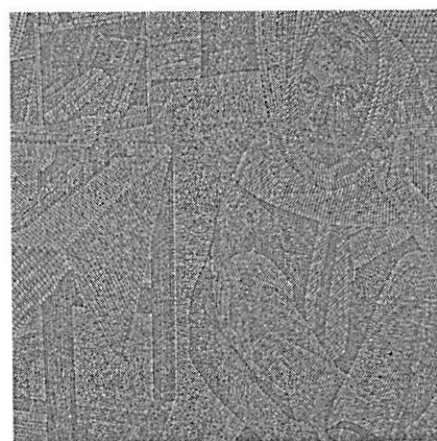
(a) Lena の 2D DTCWT 絶対値



(b) Barbara の 2D DTCWT 絶対値



(c) Lena の位相再構成画像



(d) Barbar の位相再構成画像

図 3-7: 自然画像に対する 2 次元並列木複素ウェーブレット変換

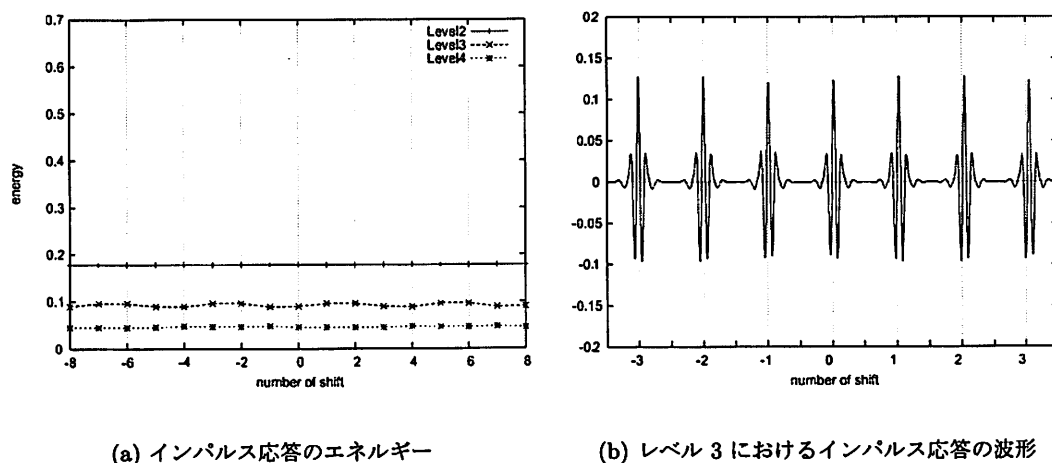


図 3-8: 並列木複素ウェーブレット変換におけるシフト不変性

DWT のシフト不変性欠如の原因はダウンサンプリングに伴うエイリアシングの発生である。ダウンサンプリングは冗長性を低減する効果があるため、シフト不変性を成立するためには冗長性を増加させなければならない。第 3.5.1 節で述べたように DTCWT ではローパス信号を 2:1 にオーバーサンプリングすることで、冗長性の増加と引き換えに近似的なシフト不変性を実現している。このため、シフト不変性に関してはフィルタバンクの制約にかかわらず、レベル 1 においても成立する。

3.6.2 高性能な方向分離特性

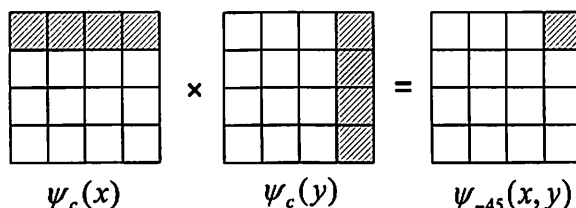
DWT は 1 次元では信号の特異点を効率的に表現できる。しかし、分離型フィルタを用いて実現される 2D DWT は画像信号の“点の特異点”は効率的に表現することができるが、エッジのような“曲線の特異点”を効率的に表現することができない。そこで、DWT よりもエッジを高い精度で近似する方向分離特性をもつ 2 次元多重解像度変換がいくつか提案されている。たとえば Steerable pyramids⁽³⁹⁾⁽⁴⁰⁾、方向性フィルタバンク⁽⁴¹⁾⁽⁴²⁾、curvelet 変換⁽⁴³⁾ などがある。これらの変換は異なる方向のエッジを異なるサブバンドへ分離する特性がある。

これらの変換と同様に 2D DTCWT も異なる方向のエッジを異なるサブバンドへ分離する性質がある。たとえば、2 次元における複素ウェーブレット $\psi_{-45}(x, y)$ は横方向の複素ウェーブレット $\psi_c(x)$ と縦方向の複素ウェーブレット $\psi_c(y)$ の積として $\psi_{-45}(x, y) = \psi_c(x)\psi_c(y)$ となる。ここで $\psi_c(x) = \psi_h(x) + j\psi_g(x)$ より $\psi_{-45}(x, y)$ は次式で与えられる。

$$\begin{aligned}\psi_{-45^\circ}(x, y) &= \psi_c(x)\psi_c(y) = [\psi_h(x) + j\psi_g(x)][\psi_h(y) + j\psi_g(y)] \\ &= \psi_h(x)\psi_h(y) - \psi_g(x)\psi_g(y) + j[\psi_g(x)\psi_h(y) + \psi_h(x)\psi_g(y)]\end{aligned}\quad (3.30)$$

1D DTCWT の周波数スペクトルはフーリエ変換と同様に片側のみ発生するため、横方向

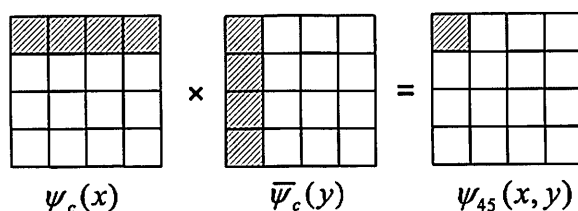
と縦方向の積は右上の一角にのみ発生する。この領域は -45° 方向のエッジに含まれる周波数帯である。



また、同様にして横方向の複素ウェーブレット $\psi_c(x)$ と縦方向の複素ウェーブレットの共役複素数 $\overline{\psi_c(y)}$ の積として与えられる2次元の複素ウェーブレット $\psi_{45}(x, y)$ は次式となる。

$$\begin{aligned}\psi_{45}(x, y) &= \psi_c(x) \overline{\psi_c(y)} = [\psi_h(x) + j\psi_g(x)][\overline{\psi_h(y) + j\psi_g(y)}] \\ &= \psi_h(x)\psi_h(y) + \psi_g(x)\psi_g(y) + j[\psi_g(x)\psi_h(y) - \psi_h(x)\psi_g(y)]\end{aligned}\quad (3.31)$$

この場合の周波数スペクトルは左上の一角にのみ発生し $+45^\circ$ 方向のエッジに含まれる周波数帯である。



式(3.30)と式(3.31)において、実部と虚部の項は2つのウェーブレット(スケーリング関数)積の加算または減算として与えられている。このときの加算処理、減算処理が第3.5.4節で述べた4つのHL, LH, HHサブバンド間の加算処理、減算処理に対応している。

横方向複素ウェーブレット(複素スケーリング関数)と縦方向複素ウェーブレット(複素スケーリング関数)の組み合わせは、このほかに $\varphi_c(x)\psi_c(y)$, $\psi_c(x)\varphi_c(y)$, $\varphi_c(x)\overline{\psi_c(y)}$, $\psi_c(x)\overline{\varphi_c(y)}$ があり、それぞれ -75° , -15° , $+15^\circ$, $+75^\circ$ 方向のエッジに含まれる周波数帯を分離する。以上より2D DTCWTでは図3-9(b)のような2次元周波数の帯域分割がなされる。図2-14(c)は全方向のエッジを保持する同心円画像(図3-9(a))に対する2D DTCWT係数の絶対値の方向分離特性である。2D DWTが 0° , 90° , $\pm 45^\circ$ の3方向のエッジしか分離できず、かつ $\pm 45^\circ$ 方向のエッジを同一サブバンドに含んでしまうのに対して、2D DTCWTでは $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ の6方向のエッジを異なるサブバンドに分離することができ、エッジなどの2次元的な特異点の解析に有効である。

3.7 並列木複素ウェーブレット係数の冗長性

2D DTCWTは4:1の冗長性を持つ冗長変換である。すなわち、 n 個の入力データを $4n$ 個の係数で表現している。このため、 $4n$ 個のDTCWT係数には多くの冗長性が存在してい

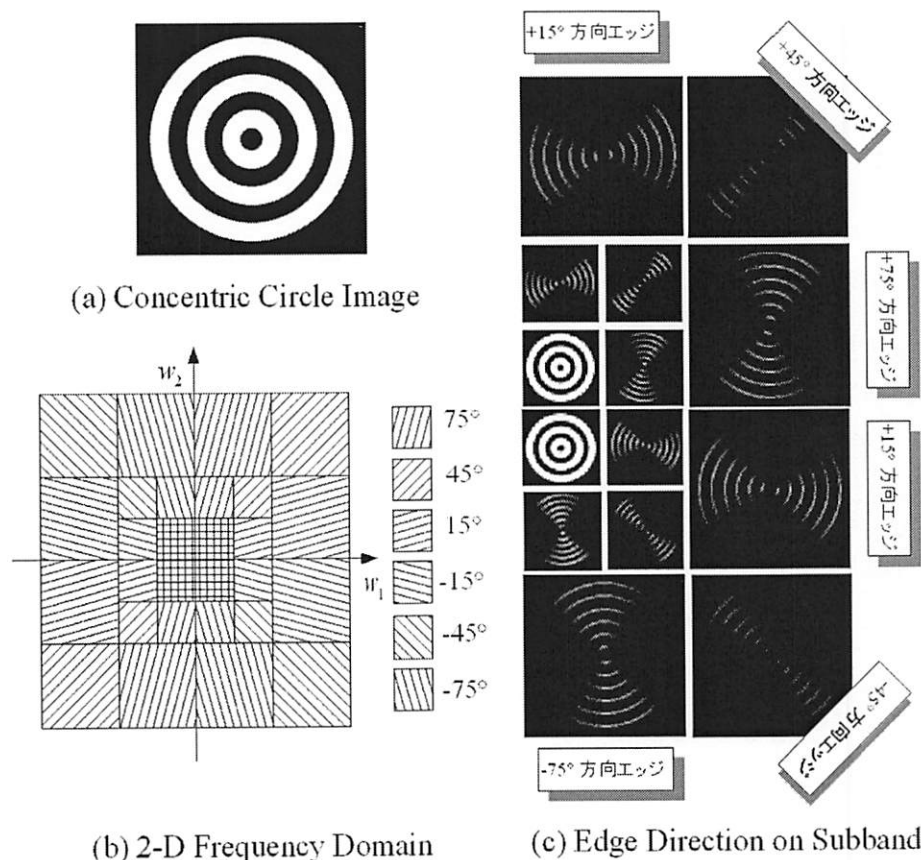


図 3-9: 2次元並列木複素ウェーブレット変換の方向分離特性

ると予想される。本節では、DTCWT 係数の冗長性について絶対値と位相に分けて述べる。

3.7.1 複素係数の絶対値の冗長性

DTCWT 係数の絶対値はシフト不変性が成り立つため、サブバンド内およびスケール間の係数に多くの冗長性が残っている。この様子を相互情報量 (mutual information) を用いて確認する。相互情報量はある変数が他の変数に関してどの程度の情報を与えるかをはかるための指標である。2 変数を $x \in X$, $y \in Y$ とし、統合確率密度関数 (joint probability density function) $p(x, y)$ に従うとする。このとき、 x と y の間の相互情報量は式 (3.32) により定義される。

$$I(X; Y) = \int_X \int_Y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (3.32)$$

ここで、 $\log$ の底は 2 である。相互情報量は X と Y に関して対称であり、 x と y が独立である場合は $I(X; Y) = 0$ であり、 x が y の関数である場合は $I(X; Y) = \infty$ となる。すなわち、相互情報量の値が大きいほど 2 変数間の冗長性が高い。2 次元のウェーブレット変換において 3 種類の相互情報量を定義することができる。

$I(X; PX)$

任意の1係数 x と同空間位置の1レベル低域の係数である親 (parent) px との相互情報量である。スケール間の冗長性の指標となる。

 $I(X; NX)$

任意の1係数 x と同一サブバンド内の8近傍である近所 (neighbors) nx との相互情報量である。サブバンド内の冗長性の指標となる。

 $I(X; CX)$

任意の1係数 x と同空間位置、同レベルの他方向サブバンドである親類 (cousins) cx との相互情報量である。サブバンド間の冗長性の指標となる。

$I(X; NX)$ と $I(X; CX)$ は1変数と多変数の相互情報量を求めている。この場合、相互情報量を求めるために必要な統合確率密度関数の次元数が $I(X; NX)$ では9次元、 $I(X; CX)$ では7次元と高くなり、これを正確に推定するために必要なデータ数が不足する。そこで、推定精度を十分保ったまま次元数を下げる必要がある。ここでは、式(3.33)のスカラ関数を用いて、 $I(X; NX)$ の代わりに $I(X; T_{NX})$ 、 $I(X; CX)$ の代わりに $I(X; T_{CX})$ を推定する(44)(45)。

$$T_{NX} = \sum_i W_i |nx_i|^2, \quad T_{CX} = \sum_i W_i |cx_i|^2 \quad (3.33)$$

ここで、 W_i は重み係数であり、参照係数の個数の逆数を取り $T_{NX} = 1/8$ 、 $T_{CX} = 1/5$ である。

Lena (512 × 512, 8[bit] グレースケール) と Barbara (512 × 512, 8[bit] グレースケール) に対する DTCWT 係数の絶対値と DWT 係数の相互情報量を表3-2から表3-5に示す。DTCWT の絶対値と DWT の相互情報量を比較すると、 $I(X; PX)$ 、 $I(X; NX)$ 、 $I(X; CX)$ のいずれについても DTCWT の相互情報量が高いことがわかる。これより、DTCWT 係数の絶対値には高い冗長性が残っている(46)。

3.7.2 複素係数の位相の冗長性

第3.4節において、式(3.5)より時間表現におけるシフトは FT における位相の回転に相当することを述べた。具体的には、時間表現におけるシフト t_0 は、FT では周波数 ω_0 において $\omega_0 t_0$ の位相回転となる。これより、位相の回転量はシフト量 t_0 と対象周波数 ω_0 に依存する。DTCWT 係数も FT 係数と同様に、物体の微小シフトに対する位相回転特性を持つ。また DTCWT は DWT と同様に周波数帯域を図3-10のようにオクターブ分割する。このため、隣接レベルのサブバンド間では2倍の帯域幅の違いがある。以上より、エッジ近辺の DTCWT 係数の位相は以下の特性を持つ。

- 同一サブバンド内の隣接する2係数において、各係数とエッジの距離がそれぞれ t_0 、 t_1 であり、対応する周波数が $\omega_0 \approx \omega_1$ のとき、位相は距離 t_0 、 t_1 に比例した回転をする。

表 3-2: Lena における並列木複素ウェーブレット係数の相互情報量 (文献⁽⁴⁶⁾ 参照)

subband	Absolute of DTCWT								
	Level1			Level2			Level3		
	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)
15	0.121	0.255	0.447	0.300	0.382	0.680	0.318	0.298	0.815
45	0.089	0.198	0.265	0.278	0.287	0.600	0.302	0.348	0.800
75	0.223	0.456	0.536	0.347	0.622	0.869	0.288	0.624	0.933
-15	0.123	0.244	0.359	0.334	0.410	0.782	0.384	0.365	0.957
-45	0.123	0.221	0.309	0.316	0.344	0.719	0.381	0.382	0.957
-75	0.213	0.420	0.519	0.360	0.661	0.910	0.318	0.711	0.997

表 3-3: Lena における離散ウェーブレット係数の相互情報量 (文献⁽⁴⁶⁾ 参照)

subband	Coefficient of DWT								
	Level1			Level2			Level3		
	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)
HL	0.050	0.040	0.142	0.166	0.148	0.315	0.257	0.221	0.497
HH	0.030	0.043	0.090	0.156	0.173	0.289	0.258	0.307	0.485
LH	0.104	0.049	0.199	0.221	0.159	0.439	0.282	0.214	0.591

表 3-4: Barbara における並列木複素ウェーブレット係数の相互情報量

subband	Absolute of DTCWT								
	Level1			Level2			Level3		
	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)
15	0.104	0.262	0.661	0.202	0.292	0.813	0.291	0.224	0.775
45	0.139	0.381	0.649	0.148	0.250	1.058	0.209	0.158	0.710
75	0.250	0.718	1.170	0.167	0.499	1.009	0.293	0.421	0.795
-15	0.097	0.244	0.686	0.222	0.291	0.861	0.354	0.256	0.856
-45	0.171	0.390	0.651	0.143	0.294	0.954	0.281	0.233	0.818
-75	0.220	0.742	1.082	0.207	0.437	0.987	0.282	0.383	0.849

表 3-5: Barbara における離散ウェーブレット係数の相互情報量

subband	Coefficient of DWT								
	Level1			Level2			Level3		
	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)
HL	0.056	0.071	0.285	0.105	0.100	0.405	0.216	0.096	0.420
HH	0.098	0.134	0.278	0.091	0.186	0.599	0.177	0.162	0.431
LH	0.160	0.179	0.620	0.112	0.144	0.546	0.255	0.106	0.444

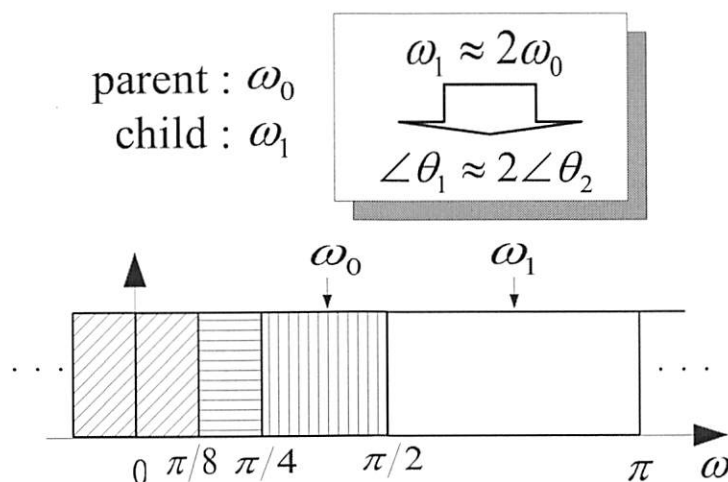


図 3-10: 並列木複素ウェーブレット変換の帯域分割特性

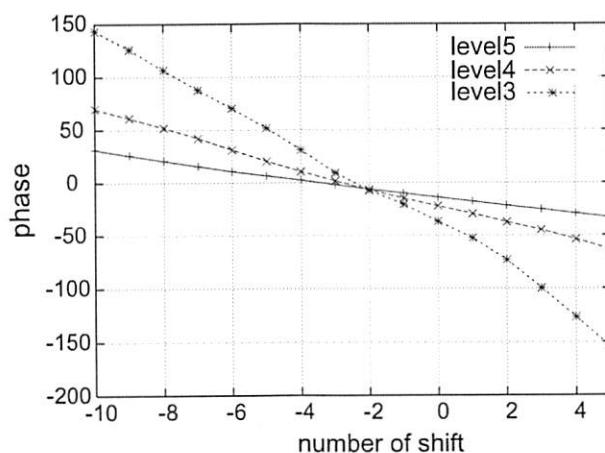


図 3-11: 並列木複素ウェーブレット変換の位相変化

- 隣接するレベル間の同位置の 2 係数において、2 係数とエッジの距離は $t_0 \approx t_1$ であり、低解像度の周波数 ω_0 は高解像度の周波数 ω_1 の半分となるため、位相の回転も半分となる (図 3-10)。

これらの特性を図 3-11 に示す。図 3-11 は 1 次元において、ある DTCWT 係数近辺をステップエッジがシフトしながら通過した際の DTCWT 係数の位相変化を示している。横軸はステップエッジのシフト量、縦軸は位相である。同一サブバンド内では DTCWT 係数とステップエッジの距離に応じて位相が線形に変化している。また、隣接レベル間では高解像度の位相変化は低解像度の位相変化の約 2 倍の傾きを持っている。このような DTCWT 係数の位相特性は、自然画像のモデル化やオブジェクト認識に利用されている (47)(48)(49)(50)。同一サブバンド内やレベル間における予測可能な位相変化は、画像符号化において位相間の冗長性であると考えられる。

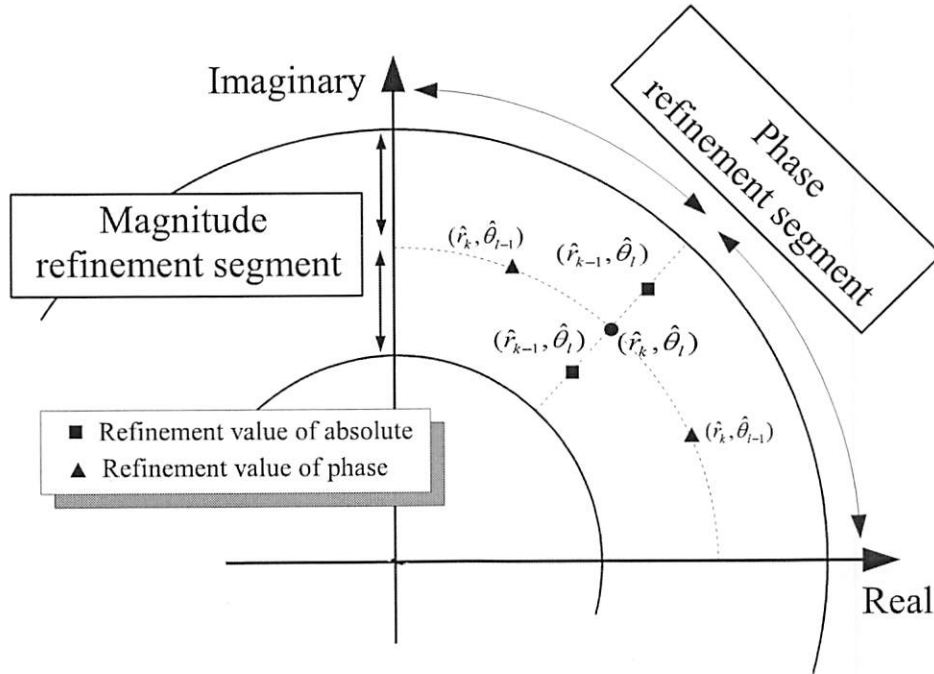


図 3-12: 複素数空間の量子化セル

3.7.3 並列木複素ウェーブレット係数の SPIHT 符号化と符号化性能

本節では、オリジナルの DTCWT 係数を用いた画像符号化の性能を確認する。Reeves らは 2D DTCWT におけるゼロツリー符号化を提案している⁽⁵¹⁾。この手法では複素数を絶対値と位相に分離して考える。まず、絶対値において第 2.5 節と同様の Sorting Pass を実行する。続いて Refinement Pass において Sortgin Pass で有意係数と判定された複素係数の精度を向上させる。この点は DWT の SPIHT 符号化と同じである。DTCWT では Refinement Pass の実行において、複素係数の精度が絶対値と位相の精度に依存する。そのため、絶対値と位相のどちらの要素が複素数の精度向上への貢献度が高いかを考慮する必要がある。Reeves らは R-D (Rate-Distortion) の観点から各要素の貢献度を判定している。

極形式による複素係数の真値を $x = re^{j\theta}$ 、量子化値を $\hat{x}_{k,l} = \hat{r}_k e^{j\hat{\theta}_l}$ とする。このとき両複素係数の 2 乗誤差は式 (3-34) となる。

$$D_x(r, \theta; \hat{r}_k, \hat{\theta}_l) = (\hat{r}_k \cos \hat{\theta}_l - r \cos \theta)^2 + (\hat{r}_k \sin \hat{\theta}_l - r \sin \theta)^2 \quad (3-34)$$

式 (3-34) に基づき絶対値の精度向上に 1[bit] を費やした場合と位相の精度向上に 1[bit] を費やした場合の誤差低減量を式 (3-35)、式 (3-36) によりあらわせる。

$$\Delta D_{k-1,l} = D_x(r, \theta; \hat{r}_k, \hat{\theta}_l) - D_x(r, \theta; \hat{r}_{k-1}, \hat{\theta}_l) \quad (3-35)$$

$$\Delta D_{k,l-1} = D_x(r, \theta; \hat{r}_k, \hat{\theta}_l) - D_x(r, \theta; \hat{r}_k, \hat{\theta}_{l-1}) \quad (3-36)$$

Refinement Pass において絶対値と位相のいずれかに 1[bit] 費やすことは、図 3-12 のように複素空間の量子化セルの領域を 4 分割したうちの 1 領域に制限することである。エンコー

ダでは式 (3.35) と式 (3.36) を比較することにより、絶対値と位相のどちらに 1[bit] を費やすべきかを決定できる。しかし、デコーダではエンコーダのパスを付加情報として受信しなければエンコーダと同様のパスを追跡できない。このようなオーバーヘッド情報は好ましくない⁽⁵²⁾。そこで、エンコーダとデコーダで共有できる歪の期待値を計算することで、同じパスを追跡している。この歪の期待値を用いて式 (3.37) が成立する限り位相の精度向上にビットを費やす。

$$E[\Delta D_{k,l-1}] > E[\Delta D_{k-1,l}] \quad (3.37)$$

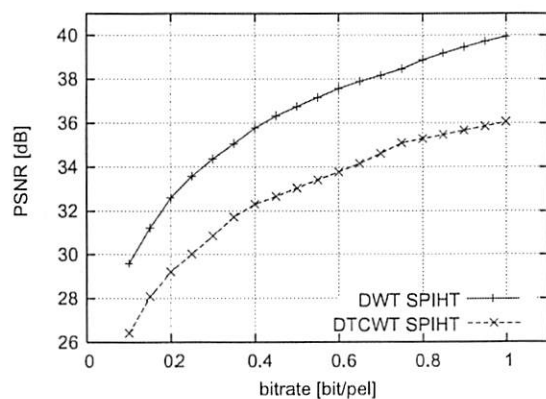
標準画像 Lena (512 × 512 サイズ, 8 [bit] グレースケール), Barbara (512 × 512 サイズ, 8 [bit] グレースケール) および Goldhill (512 × 512 サイズ, 8 [bit] グレースケール) に対する DTCWT 係数の SPIHT 符号化の性能を図 3-13 に示す。図 2-18(a)(c)(e) はレート歪み特性であり、DWT の SPIHT 符号化に対して、DTCWT の SPIHT 符号化は PSNR が約 2.0[dB] – 4.0[dB] 低く、性能が悪いことがわかる。また、図 2-18(b) は 0.2 [bit/pel] における Lena の再構成画像、図 2-18(d) は 0.5 [bit/pel] における Barbara の再構成画像、図 2-18(f) は 0.3 [bit/pel] における Goldhill の再構成画像である。DWT のアーティファクトと異なり、DTCWT のアーティファクトはエッジに沿って発生している。

DTCWT のような冗長変換では、多数の変換係数により冗長に入力信号を表現しているため、これらの変換係数を単純に符号化しただけでは高い符号化効率が得られない。そこで、次節で述べる疎表現と呼ばれる係数列に対して符号化を行う必要がある。

3.8 冗長変換の疎表現

3.8.1 冗長変換と疎表現の概念

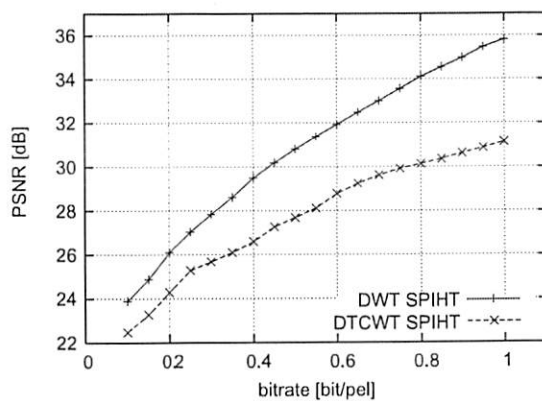
非冗長変換は完備な基底を用いる変換であり、変換前後におけるデータ数は不変である。一方で、過完備な基底を用いた冗長変換は、変換後のデータ数が増加する変換である。しかし、冗長変換において、疎表現と呼ばれる表現を求めることにより、非冗長変換よりも少ない非ゼロ係数 (基底) で信号を表現することができる。この様子を図 3-14 に示す。図 3-14 は $\mathbb{R}^3$ の信号 x を基底の線形結合により表現している例である。直交基底 $\mathcal{D}_1 = \{\phi_1 \ \phi_2 \ \phi_3\}$ を用いた (a) の場合では、信号 x を 3 つの基底により $x = \alpha_1\phi_1 + \alpha_2\phi_2 + \alpha_3\phi_3$ と表現するしかない。一方で、(b) の場合は、直交基底 $\mathcal{D}_1$ とそれらを原点周りに回転してできた別の直交基底 $\mathcal{D}_2 = \{\psi_1 \ \psi_2 \ \psi_3\}$ から構成される過完備な基底集合 $\mathcal{D} = \{\mathcal{D}_1 \ \mathcal{D}_2\}$ により信号 x を表現する。今回の例では、信号 x は基底 ψ_2, ψ_3 の張る平面上にあるため、 $x = \alpha_1\psi_2 + \alpha_2\psi_3$ と表現することが可能である。すなわち、(b) の場合では (a) の場合よりも信号表現に用いる基底が 1 つ少なくなっており、非ゼロ係数の個数も 1 つ少ない。疎表現とは、このように信号の特性を良くあらわす基底の集合を用意し、信号表現に用いる基底 (非ゼロ係数) の個数を少なくしたものである。



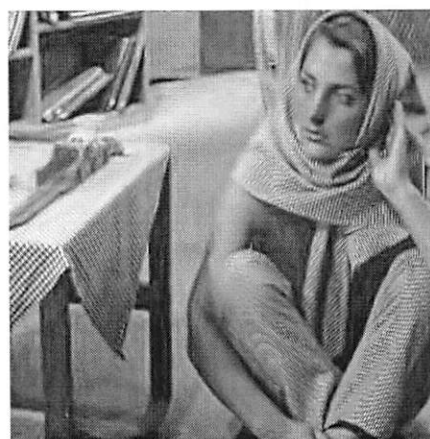
(a) Lena のレート歪み特性



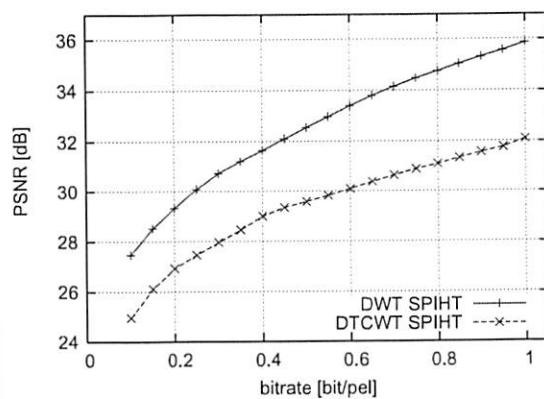
(b) 0.2 (bit/pel) : 29.21 (dB)



(c) Barbara のレート歪み特性



(d) 0.5 (bit/pel) : 27.69 (dB)



(e) Goldhill のレート歪み特性



(f) 0.3 (bit/pel) : 27.97 (dB)

図 3-13: 並列木複素ウェーブレット係数に対する SPIHT 符号化の性能

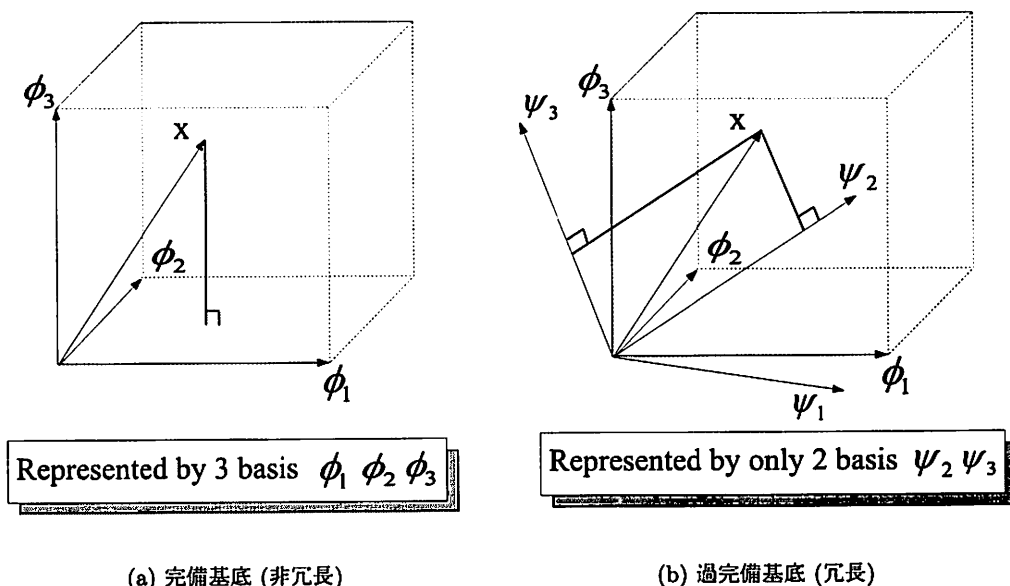


図 3-14: 基底による信号表現

3.8.2 疎表現の理論

n 次元のベクトル $y \in \mathbb{R}^n$ を基底の集合 $\mathcal{D} = \{\phi_i \in \mathbb{R}^n, i = 1, \dots, m\}$ を用いて線形結合により表現する問題を考える. 一般的に基底の集合 $\mathcal{D}$ は辞書 (dictionary), 辞書の各基底 ϕ_i は原子 (atoms) と呼ばれる⁽⁵⁸⁾.

$$y = \sum_i \alpha_i \phi_i \quad (3.38)$$

$n = m$ であり, 各基底が直交基底であるとき, 辞書 $\mathcal{D}$ は完備な基底から構成されている. このとき, 式 (3.38) は線形独立であり, この方程式の解 α_i はただひとつしか存在しない. すなわち, y を表現する基底の組み合わせは 1 通りである. JPEG や JPEG 2000 で用いられている DCT や DWT は直交変換であり, この理論が当てはまる. 一方で, $m > n$ となるとき基底の数が原信号の次元より多いため, 辞書 $\mathcal{D}$ は過完備な基底から構成されている. このとき, 式 (3.38) は必ず線形従属となり, この方程式の解 α_i は無数に存在する. すなわち, y を表現する基底の組み合わせが無数に存在する. この無数に存在する表現の中で, 疎表現とは最少数の基底を用いて原信号 y を表現したものである. すなわち, 辞書に対応する $n \times m$ 行列 D を用いて, 疎表現を求める問題は以下のベクトル $\alpha \in \mathbb{R}^m$ を求めることである.

$$\text{Minimize } \|\alpha\|_0 \quad \text{subject to } D\alpha = y \quad (3.39)$$

ここで, $\|\alpha\|_0$ は α の l^0 ノルムであり, 非ゼロ係数の個数を表している.

式 (3.39) は厳密に原信号 y を表現する疎表現を求めているが, 現実的にはある程度の誤差を許容できる場合が多い. このとき, 既知の信号 y が原信号 s と白色雑音 z から構成さ

れているとする.

$$\mathbf{y} = \mathbf{s} + \sigma \mathbf{z} \quad (3.40)$$

ここで, $\sigma > 0$ は雑音レベルのパラメータである. この場合, $\mathbf{y}$ を表現するのではなく, $\mathbf{s}$ を表現する必要がある. そこで, 誤差を含む場合の疎表現の問題は式 (3.41) となる.

$$\text{Minimize } \|\alpha\|_0 + \lambda \|\mathbf{D}\alpha - \mathbf{y}\|_2^2 \quad (3.41)$$

解 α はパラメータ λ の関数であり, $\mathbf{y}$ は式 (3.42) により表現される.

$$\mathbf{y} = \mathbf{s}^{(\lambda)} + \mathbf{r}^{(\lambda)} \quad (3.42)$$

ここで, $\mathbf{s}^{(\lambda)} = \mathbf{D}\alpha^{(\lambda)}$ であり, $\mathbf{r}^{(\lambda)}$ は誤差である. 誤差の大きさは λ により制御される. たとえば, $\lambda \rightarrow \infty$ とすると式 (3.41) の第二項の影響が強くなるため, $\mathbf{D}\alpha \approx \mathbf{y}$ となる解 α が求まる. すなわち, $\mathbf{r}^{(\lambda)} \rightarrow 0$, $\mathbf{s}^{(\lambda)} \rightarrow \mathbf{y}$ となる. 一方で, $\lambda \rightarrow 0$ とすると式 (3.41) の第二項の影響が弱くなるため, $\mathbf{D}\alpha$ と $\mathbf{y}$ の誤差の影響が小さくなる. よって, $\mathbf{r}^{(\lambda)} \rightarrow \mathbf{y}$, $\mathbf{s}^{(\lambda)} \rightarrow 0$ となる. 疎表現係数列は最少数の非ゼロ係数で信号を表現するため, 1 係数当たりのエネルギー集約度が非常に高い係数列である. 少数の係数にエネルギーが集中すれば, 信号復元に必要な有意係数が減少するため, データ圧縮に望ましい.

なお, 2D DTCWT は $m = 4n$ であり, 過完備な基底の集合である. TreeA*TreeA の辞書を Φ_{aa} , TreeA*TreeB の辞書を Φ_{ab} , TreeB*TreeA の辞書を Φ_{ba} , TreeB*TreeB の辞書を Φ_{bb} とする. 各辞書は $n \times n$ の近似的な直交基底から構成されており, 2D DTCWT の辞書は $\mathbf{D} = \{\Phi_{aa} \Phi_{ab} \Phi_{ba} \Phi_{bb}\}$ となる⁽⁵³⁾.

式 (3.39), 式 (3.41) の問題は NP (Non-deterministic Polynomial time) 困難であり, その解を求めるには α の全要素の組み合わせを探索する必要があり, 計算が複雑である. そこで, この解を近似的に求める手法がいくつか提案されている. ここでは, その中でも有名はマッチング追跡 (MP : Matching Pursuit)⁽⁵⁴⁾⁽⁵⁵⁾ と基底追跡 (BP : Basis Pursuit)⁽⁵⁸⁾ について述べる.

3.8.3 マッチング追跡近似解法

MP は反復処理に基づくアルゴリズムである. 1 回のステップで, 原信号と近似信号の 2 乗誤差を最小にするように辞書 $\mathcal{D}$ から原子を 1 つ選択する. ここでは, MP に直交化のステップを付加した直交マッチング追跡 (OMP : Orthogonal Matching Pursuit) について述べる⁽⁵⁵⁾. MP は貪欲アルゴリズム (greedy algorithm) と呼ばれ, 各ステップにおいて 2 乗誤差を最も小さくする基底を選択することで, 最終的に得られる結果の 2 乗誤差も最小となる, という考え方に基づいている.

MP の初期状態は, $\hat{\mathbf{r}}^{(0)} = \mathbf{y}$ であり, 残差信号は原信号となっている. また, 近似信号は $\hat{\mathbf{y}}^{(0)} = 0$ である. この初期状態より, 段階的に基底を選択することで疎表現の近似を行う. 1 ステップの処理において, 近似信号は $\hat{\mathbf{y}}^{(k-1)} \rightarrow \hat{\mathbf{y}}^{(k)}$, 残差信号は $\hat{\mathbf{r}}^{(k-1)} \rightarrow \hat{\mathbf{r}}^{(k)}$ にそれぞれアップデートされる. このとき, 常に $\mathbf{y} = \hat{\mathbf{y}}^{(k)} + \hat{\mathbf{r}}^{(k)}$ が成立している. ここで, $\|\phi_i\|_2 = 1$

と正規化されているとする。 k 回目のステップで、辞書 $\mathcal{D}$ の中から残差信号との相関値が最も大きい原子を 1 つ選択し、近似信号に追加する。

$$i_k = \operatorname{argmin}_{1 \leq i \leq m} |\langle \hat{r}^{(k-1)} | \phi_i \rangle| \quad (3.43)$$

k 回目のステップまでに選ばれた原子を用いて、近似信号は式 (3.44) となる。

$$\hat{y}^{(k)} = \sum_{i=1}^k a_{i_i}^k \phi_{i_i} \quad (3.44)$$

係数 $a_{i_i}^k$ は原信号と近似信号の 2 乗誤差 $\|y - \hat{y}^{(k)}\|^2$ を最小にするよう決定される。残差信号は原信号から式 (3.45) の近似信号を差分したものとして再定義される。

$$r^{(k)} = y - \hat{y}^{(k)} \quad (3.45)$$

このとき、残差信号 $r^{(k)}$ は k ステップまでに選ばれたすべての基底に直交していると考えることができる。この残差信号が $k+1$ 回目のステップの入力となり同様の処理が繰り返される。MP (OMP) は残差信号のノルムがあらかじめ定義した閾値を下回った場合や所望の原子数を選択した場合に、処理を終了させることができる。そして、その結果が原信号に対する疎表現となっている。

辞書 $\mathcal{D}$ が直交基底から構成されている場合、MP は最適な疎表現を与える。しかし、辞書が過完備基底のように直交基底から構成されていない場合、MP を用いた疎表現近似手法は失敗する場合がある。疎表現の近似手法に MP を用いる場合の欠点は、アルゴリズムの貪欲さである。MP は各ステップで局所的に最適と判断される原子を選択するが、それらが大局的にみて最適であるとは限らない場合がある。このような場合では、初期の数ステップで本来の疎表現には含まれない原子を選んでしまう。その結果、その後のステップで選ばれた原子も誤ったものとなる可能性が高くなる。これに対して BP はこのような場合にも良好な疎表現を求めることができる。

3.8.4 基底追跡近似解法

BP は式 (3.39) の l^0 ノルムを l^1 ノルムに置き換え、 α の l^1 ノルムを最小にする。

$$\text{Minimize } \|\alpha\|_1 \quad \text{subject to } D\alpha = y \quad (3.46)$$

また、式 (3.41) に対応する誤差を含む場合の BP は式 (3.47) であり、基底追跡ノイズ除去 (BPDN : Basis Pursuit De-Noising) と呼ばれる。

$$\text{Minimize } \|\alpha\|_1 + \lambda \|D\alpha - y\|_2^2 \quad (3.47)$$

l^1 ノルムは凸であるため、BP (BPDN) は凸最適化問題であり、線形計画法 (LP : Linear Programming) を用いて解を求めることができる。BP は MP よりも良好な疎表現を与える場合があることが知られている。

LP は次のように定義される。行列 $A \in \mathbb{R}^{m \times n}$, 列ベクトル $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ が与えられたとき, 式 (3.48) を満足する $x \in \mathbb{R}^n$ を求める。

$$\text{Minimize } c^T x \quad \text{subject to } Ax = b, \quad x_i \geq 0 \quad (3.48)$$

式 (3.46) は次の変数の置き換えにより式 (3.48) の LP となる。

$$\begin{aligned} A &\leftarrow [\phi_1 \ \phi_2 \ \cdots \ \phi_n \ -\phi_1 \ -\phi_2 \ \cdots \ -\phi_n] \\ b &\leftarrow x \\ c &\leftarrow [1 \ 1 \ \cdots \ 1] \\ x &\leftarrow [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_n] \end{aligned}$$

式 (3.48) において, $c^T x$ を最小にすることは, 式 (3.46) における l^1 ノルムを最小にすることと等価である。A は辞書 D に対応する行列であるが, LP では係数は非負であるという制約があるため, 各原子の負成分も含め 2 倍に拡張されている。

LP のアルゴリズムは種々あるが, BP では単体アルゴリズム (simplex algorithm) と内点アルゴリズム (interior-point algorithm) が適用できることが知られている⁽⁵⁸⁾。

単体アルゴリズム LP における単体アルゴリズムでは, まず, 行列 A の n 個の線形独立な列ベクトルから構成される初期行列 B を見つける。このときの解 $B^{-1}b$ は実行可能解 (feasible solution) である。次に, 反復処理を用いて, 各反復において目的関数を最大にするように, 現在の基底集合 B に属する基底と属さない基底を入れ替える。最適解でない場合には, 常に目的関数を改善するか維持する入れ替えが存在する。

BP の場合では, まず, 辞書 D から線形独立な n 個の原子を選び, これを初期の原子集合とする。次に, 各反復において, 目的関数を改善するために, 原子集合に含まれる原子と含まれない原子を入れ替える。改善が不能になるまで繰り返し基底の改善を行う。

内点アルゴリズム 実行可能な点の集合 $\{x : Ax = b, x \geq 0\}$ は $\mathbb{R}^n$ の凸多面体 (convex polyhedron), すなわち, 単体 (simplex) である。幾何的には, 単体法では, この単体の境界を探索し, 多面体のある頂点から目的関数を改善する別の頂点へ移動する。これに対して, 内点法では単体の内部 $x^{(0)}$ から開始し, 反復処理により $x^{(k)}$ は境界へ近づいていく。

BP の場合では, 解 $D\alpha^{(0)} = s$ with $\alpha^{(0)} > 0$ から開始する。次に, 各反復において, $D\alpha^{(k)} = s$ を維持しながら, ベクトル $\alpha^{(k)}$ が疎にするような変形を行う。反復過程において, $\alpha^{(k)}$ は n よりも少ない非ゼロ係数を持つようになり, その係数に対応する原子が最終的な解を表現する可能性が高くなる。

BP の幾何的解釈は次のとおりである。 $\mathbb{R}^d$ 次元空間において, 信号はベクトル, 原子は 1 点として扱うとする。また, 辞書 D は $n(> d)$ 個の原子からなるとする。このとき, $\|\alpha\| = 1$

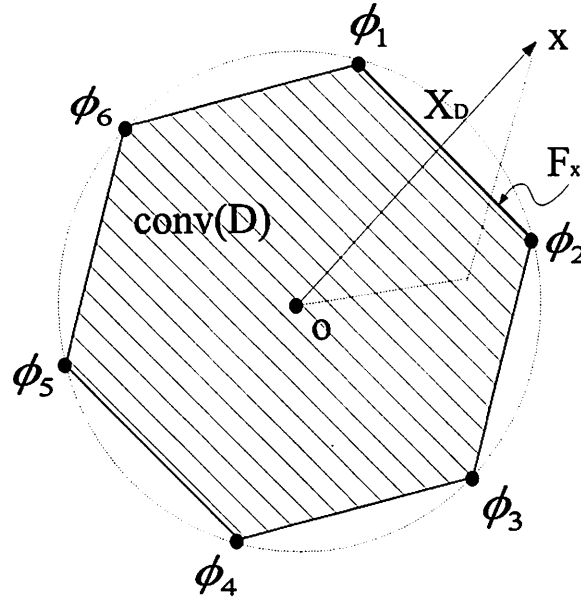


図 3-15: 基底追跡の幾何的解釈

となる α により表現される信号の集合を考える。この集合は辞書 $\mathcal{D}$ の凸包 (convex hull) である。

$$\text{conv}(\mathcal{D}) = \{x | x = \sum_{i \in \Gamma} \alpha_i \phi_i \text{ and } \sum_{i \in \Gamma} \alpha_i = 1, \alpha_i > 0\} \quad (3-49)$$

ここで、 $\Gamma \subseteq 1, \dots, n$ である。 $\|\phi_i\|_2 = 1$ とすると、 $\text{conv}(\mathcal{D})$ は単位球に内接する多面体である。この様子を図 3-15 に示す。この例では $n=6$ である。ベクトル x と $\text{conv}(\mathcal{D})$ の境界との交点を x_D とすると、 x_D は面 F_x の頂点の線形結合により表現することができる。この例では、 x を表現する原子は ϕ_1, ϕ_2 となる。このときの表現は最小の l^1 ノルムを与え、 $l^1 = 1$ である。すなわち、ベクトル x により通過される $\text{conv}(\mathcal{D})$ の面の頂点が、 l^1 ノルムの最小値を与える x の原子となっている。つまり、BP は x が通過する $\text{conv}(\mathcal{D})$ の面を見つけることと等価である。

3.9 むすび

本章では、DTCWT による新しい静止画像符号化について述べた。まず、2 種類の CWT があることを述べ、ヒルベルト変換について説明した。次に DTCWT について述べ、DTCWT の特性が DWT の問題点を解決することを述べた。そして、DTCWT 係数の冗長性を示し、DTCWT 係数の SPIHT 符号化の性能を確認し、疎表現の必要性を述べた。最後に、冗長変換における疎表現の理論と疎表現を求める近似解法について述べた。

第 4 章では、本章の内容に基づき、DTCWT の疎表現を用いた静止画像符号化の検討を行う。

第4章

並列木複素ウェーブレット変換の疎表現における係数選択手法と係数予測手法

4.1 まえがき

第3章において、オリジナルの DTCWT 係数では高い符号化効率が得られないことを確認し、疎表現係数列の必要性を述べた。本章では、DTCWT の疎表現係数列を用いた静止画像符号化の検討を行う。まず、本研究において想定している DTCWT の疎表現係数列を用いた符号化システムの概要を述べる。次に、DTCWT 係数に対する雑音変形手法である Noise Shaping 法を説明し、これが l^0 ノルムベース交互射影法と呼ばれる疎表現近似解法と等価であることを示す。そして、その等価性より Noise Shaping 法における新しい係数選択手法を提案する。次に、提案手法の性能と DTCWT の疎表現係数列の符号化性能を確認する実験を行い、疎表現を用いた静止画像符号化の考察を行う。最後に、DTCWT 画像符号化の符号化効率改善手法として、疎表現係数の絶対値を予測する手法を提案し、実験によりその有効性を示す。

4.2 並列木複素ウェーブレット変換の疎表現を用いた静止画像符号化システム

本研究において想定している DTCWT の疎表現係数列を用いた符号化器と復号化器を図 4-1 に示す。符号化器は次のような構成である。まず、入力画像に対して 2D DTCWT を実行する。次に、疎表現係数列を求めるため疎表現近似解法を適用する。そして、疎表現係数列に対して、量子化とエントロピー符号化を行う。本章では、疎表現近似解法の考察と疎表現近似解法における新しい係数選択手法を提案をしている。また、量子化として、スカラ量子化と単純なベクトル量子化を適用し、ベクトル量子化の優位性を言及する。しかし、疎表現係数列におけるエントロピー符号化や具体的なベクトル量子化は未検討であり今後の課題である。

復号化器では、まずエントロピー復号化と逆量子化を行う。その後、逆量子化された係数列を用いてサブバンドの予測を行う。復号化器では、エネルギーの集中した少数の非ゼロ係数のみを受信した状態となっている。そこで、これらの受信係数列より非受信の係数列を予測することを試みる。本章では、復号化器において非受信係数の絶対値のみを予測する提案

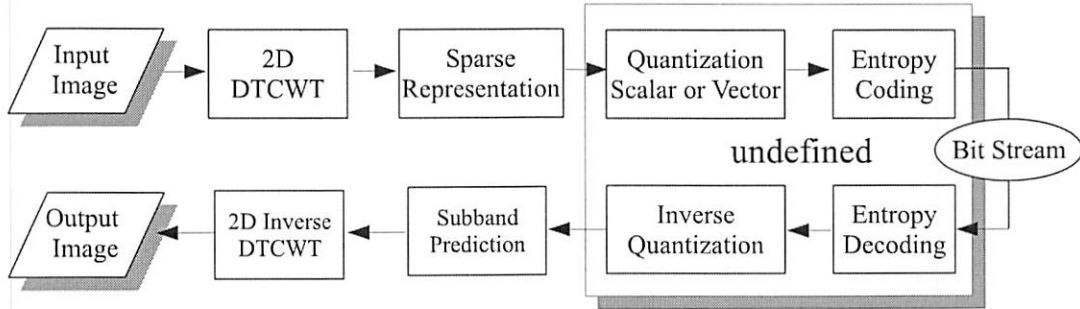


図 4-1: 並列木複素ウェーブレット変換の疎表現係数列における符号化器と復号化器

している。最後に、2D DTCWT の逆変換を行い再構成画像を復元する。

4.3 疎表現近似解法の考察と係数選択手法の提案

4.3.1 Noise Shaping 法

非最大間引きフィルタバンクは最大間引きフィルタバンクに比べ、サブバンド上で付加された雑音が再構成信号に与える影響を小さくする特性がある。この特性を利用して、反復処理により雑音を変形し、再構成信号へ与える雑音の影響をさらに小さくする手法が Noise Shaping 法である⁽⁵⁹⁾。

Kingsbury らはこの手法を DTCWT へ適用し、結果的に疎表現係数列を生成する提案をしている⁽⁶⁰⁾。図 4-2 にその構成図を示す。 $\mathbf{x}$ は N 次元ベクトルの入力信号、 $\mathbf{A}$ は $M \times N$ の分解行列 ($M > N$)、 $\mathbf{y}$ は M 次元の DTCWT 係数であり、 $\mathbf{y} = \mathbf{A}\mathbf{x}$ が成立する。また $\mathbf{R}$ は $N \times M$ の再構成行列であり、 $\mathbf{x} = \mathbf{R}\mathbf{y}$ が成立する。関数 $f(y_i, \theta_i)$ は式 (4.1) のような閾値 θ を用いた非線形処理を行う。

$$\hat{y}_r + j\hat{y}_i = \begin{cases} 0 & \text{if } |y_r + jy_i| < \theta \\ y_r + jy_i & \text{otherwise} \end{cases} \quad (4.1)$$

ここで、 y_r 、 y_i はそれぞれ複素係数の実部と虚部である。2D DTCWT では $M = 4N$ であり、再構成処理は $4N$ 次元の係数空間から N 次元の画像空間への射影となる。よって $4N$ 次元の係数空間は、 N 次元の値域 S とこれに直交する $3N$ 次元の零空間 $\perp$ から構成される。ここでは、値域に属するウェーブレット係数を $\mathbf{y}^S$ 、零空間に属するウェーブレット係数を $\mathbf{y}^\perp$ とする。また、値域と零空間の意味は次のとおりである。

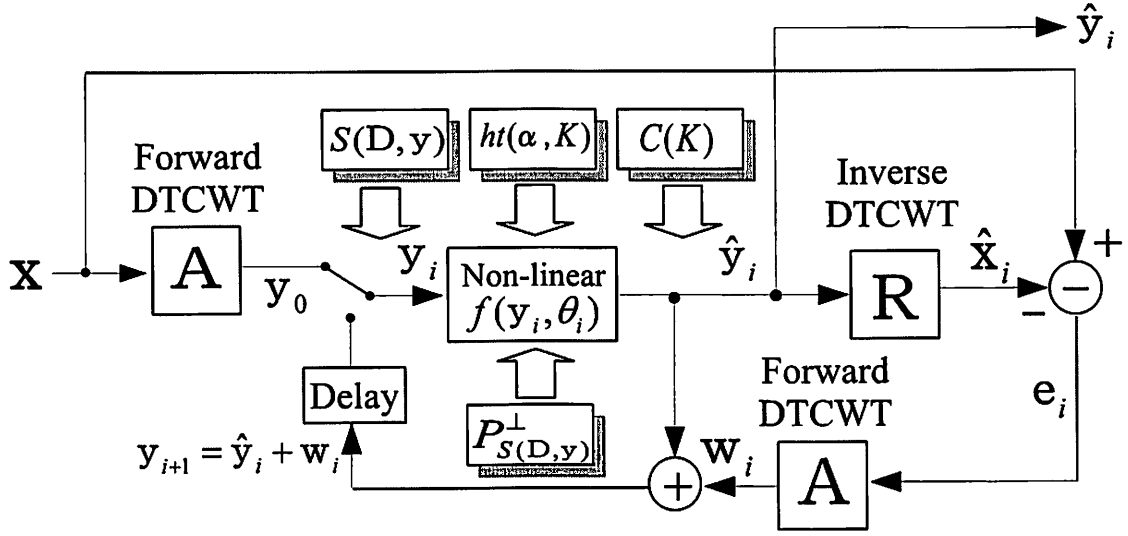
値域 再構成信号に影響を及ぼすウェーブレット係数空間。 $\mathbf{x} = \mathbf{R}\mathbf{y}^S$ が成立

零空間 再構成信号に影響を及ぼさないウェーブレット係数空間。 $\mathbf{R}\mathbf{y}^\perp = 0$ が成立

なお、値域と零空間の数学的な理論と例を付録 A に示す。

$4N$ 次元の係数空間はこれらの部分空間より式 (4.2) により構成される。

$$\mathbf{y} = \mathbf{y}^S + \mathbf{y}^\perp \quad (4.2)$$


 図 4-2: 並列木複素ウェーブレット変換における Noise Shaping 法 (文献⁽⁴⁶⁾ 参照)

ここで、式(4.1)による閾値処理は、ウェーブレット係数 y に量子化雑音 d を付加していると考えることができる。このとき雑音 d も値域と零空間から $d = d^S + d^\perp$ により構成されている。

Noise Shaping 法は反復処理を用いて、この雑音 d が再構成画像へ与える影響を小さくしてゆく。 i 回目の閾値処理後の係数は式(4.3)となる。

$$\hat{y}_i = y_i + d_i = y_i^S + y_i^\perp + d_i^S + d_i^\perp \quad (4.3)$$

雑音を付加された係数 $\hat{y}_i$ の再構成画像 $\hat{x}_i$ と入力画像 x の再構成エラーを e_i とする。この再構成エラーを再び変換してできる更新係数 w_i は式(4.4)となる。

$$w_i = Ae_i = A(x - R\hat{y}_i) = A(x - Ry_i^S - Rd_i^S) \quad (4.4)$$

ここで、 $Ry_i^\perp = 0$, $Rd_i^\perp = 0$ を用いている。雑音を付加された係数 $\hat{y}_i$ に更新係数 w_i を加算したものが、 i 回目の反復処理で得られる係数 y_{i+1} となる。

$$\begin{aligned} y_{i+1} = \hat{y}_i + w_i &= y_i + d_i + A(x - Ry_i^S - Rd_i^S) \\ &= y_i + d_i^S + d_i^\perp - ARd_i^S \\ &= y_i^S + (y_i^\perp + d_i^\perp) \end{aligned} \quad (4.5)$$

ここで、 $Ax = y_0$, $AR = I$ を用いており、 y_0 は初期のウェーブレット係数である。また、 $ARy_i^S = y_0$ である。式(4.5)より、Noise Shaping 法は零空間のウェーブレット係数 $y_i^\perp$ を零空間の量子化雑音 $d_i^\perp$ の加算により変形する処理である。変形される空間が零空間のみであるため、再構成信号は入力信号に一致する。

反復処理の過程で、閾値処理により生じた零空間に属する誤差 $d^\perp$ は次第に小さくなり $\|d_\infty^\perp\| \rightarrow 0$ と収束する。また、それに伴い再構成エラー e_i も $\|e_\infty\| \rightarrow 0$ となる。よって、収

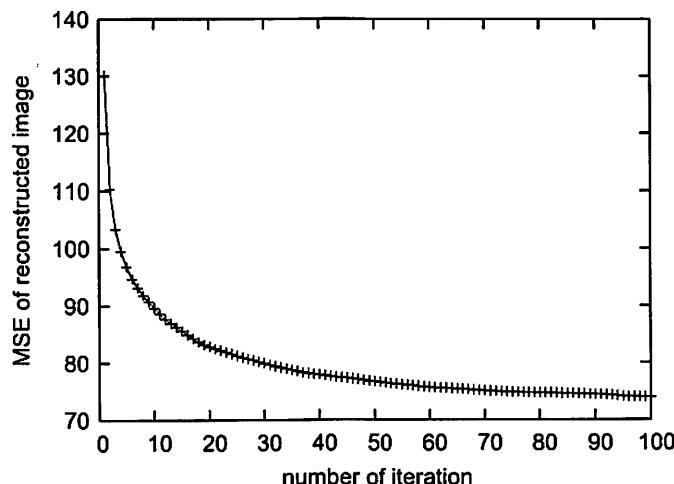


図 4-3: Noise Shaping 法の収束過程

束時では，ウェーブレット係数 y_∞ に対する式 (4.1) の閾値処理より発生する誤差 d_∞ の大部分は値域に属している．Noise Shaping の収束過程を図 4-3 に示す．図 4-3 は反復回数とともに再構成エラーが低減してゆく様子を示している．

4.3.2 Noise Shaping 法の問題点

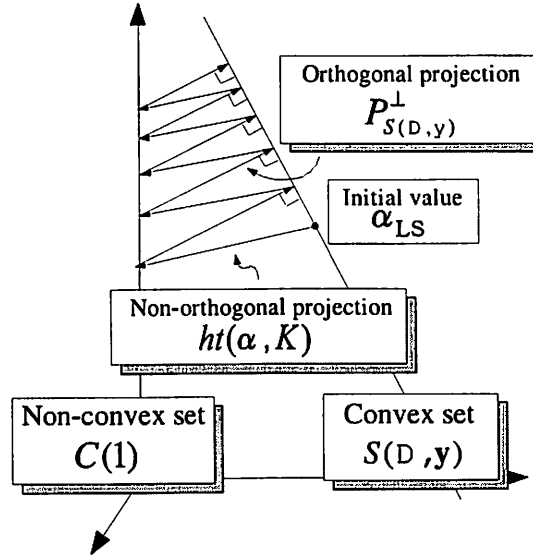
Noise Shaping 法は少数のウェーブレット係数により多くのエネルギーが集中する係数列を作り出す．つまり，結果的に Noise Shaping 法は疎表現の近似解を求める処理となっている．しかし，第 3.8.2 節で述べた疎表現問題との関係性について一切述べられていないため，疎表現の近似解が求まる理由や，その近似解の精度などが不明である．そこで，次節において Noise Shaping 法が疎表現の近似解を求める l^0 ノルムベース交互射影法と等価であることを示す．これより，疎表現問題に対する Noise Shaping 法の効果が明らかになる．

4.3.3 Noise Shaping 法の疎表現的考察

Mancera らは，式 (3.41) の雑音を含む場合の疎表現問題をある値 λ を用いて解くことが，定数 K を用いて $\|\alpha\|_0 = K$ という条件で解くことと等価であることより，式 (4.6) のように問題を再設定している．

$$\hat{\alpha}(K) = \arg \min_{\alpha \in C(K)} \|\mathbf{D}\alpha - \mathbf{y}\|_2 \quad \text{such that } C(K) = \{\alpha \in \mathbb{R}^m : \|\alpha\|_0 = K\} \quad (4.6)$$

式 (4.6) は原信号の変換係数 $\mathbf{y}$ とのユークリッド距離を最小化する K 個の非ゼロ係数からなる変換係数 α を探していると解釈できる．Mancera らは， $\hat{\alpha}(K)$ を求めるため，式 (3.38) の解集合 $S(\mathbf{D}, \mathbf{y})$ と $C(K)$ の間の交互射影を繰り返している⁽⁶¹⁾．ここで， $S(\mathbf{D}, \mathbf{y})$ は式 (3.38) の線型方程式のアフィン集合であり凸集合である．一方で， l^0 ノルムからなる $C(K)$ は凸集


 図 4-4: 交互射影による l^0 ノルム最小化 (文献⁽⁶¹⁾ 参照)

合ではない。そのため、これらの空間の射影により求まる収束点は局所解 (local minimum) である。 $C(K)$ から凸集合 $S(D, y)$ への射影は直交射影が可能で式 (4.7) であらわされる⁽⁶²⁾。

$$P_{S(D, y)}^\perp = \alpha + \alpha_{LS} - D^\dagger D \alpha \quad (4.7)$$

ここで、 $D^\dagger$ は D の擬似逆行列であり、 $D^\dagger = D^T [D D^T]^{-1}$ である。また、 $\alpha_{LS} = D^\dagger y$ は、 α の l^2 ノルムを最小にする解である。凸集合 $S(D, y)$ から $C(K)$ への射影 $ht(\alpha, K)$ は $C(K)$ と $S(D, y)$ のユークリッド距離を最小にするために、最大値から K 個の係数を選択する処理である。 $K = 1$ における反復処理の幾何的な遷移を図 4-4 に示す。Mancera らは、この手法を Curvelet 変換に適用し BP (または BPDN) よりも高いエネルギーコンパクションが得られることを示している。

第 4.3.1 節で述べた Noise Shaping 法は、ここまで述べてきた疎表現の近似解を求める l^0 ノルムベース交互射影法と等価であることを示せる⁽⁴⁶⁾。まず、Noise Shaping における式 (4.1) の閾値処理は閾値以上の係数を残す処理である。これは l^0 ノルムベース交互射影法における $ht(\alpha, K)$ に相当し、最大値から K 個の非ゼロ係数を選択していることと同じである。また、 l^0 ノルムベース交互射影法における直交射影 $P_{S(D, y)}^\perp$ は式 (4.7) は、次のように変形できる。

$$\begin{aligned} P_{S(D, y)}^\perp &= \alpha + \alpha_{LS} - D^\dagger D \alpha \\ &= \alpha + D^\dagger (y - D \alpha) \end{aligned} \quad (4.8)$$

これは Noise Shaping 法において、更新係数 w_i を雑音を付加された係数 $\hat{y}_i$ に加算する操作と同じである。これらの対応関係を図 4-2 に示す。この等価性より、疎表現を求めること

は少なくとも零空間の係数を変形することであり、疎表現の誤差のほとんどは値域から発生する、ということが言える。

この手法において $C(K)$ は凸集合でないため収束点への経路によって局所解が異なってくる。凸集合 $S(D, y)$ から $C(K)$ への射影方法には3種類考えられる。

1. $C(K)$ を固定とする。反復過程において、非ゼロ係数の選択個数が一定。
2. $C(K)$ を次第に減少させる。反復過程において、非ゼロ係数の選択個数を次第に減少。
3. $C(K)$ を次第に増加させる。反復過程において、非ゼロ係数の選択個数を次第に増加。

Kingsbury らは3. の非ゼロ係数の選択個数を次第に増加することで最も良い局所解が得られることを述べている。これは Noise Shaping 法がマッチング追跡と同様の貪欲アルゴリズムであり、近似精度の高い基底から優先的に選択することで、近似精度の悪い基底が排除を排除できるためである。ただし、Noise Shaping 法は一度選択した基底は反復終了時まで保持されるという特性があるため、Huggings らが提案⁽⁵⁶⁾ するような反復過程で不適切になった基底を除去する操作を加えることで、より良い局所解が求まる可能性があり、今後の課題である。

また、 l^0 ノルムベース交互射影法を用いた疎表現近似解法との等価性より、DTCWT において、Noise Shaping 法により求めた疎表現の近似解は、Curvelet 変換の場合と同様に BP (または BPDN) よりも近似精度の良い疎表現を求めている可能性がある。しかし、近似解法の性能は用いる辞書によって変わるため、BP (または BPDN) との厳密な比較は DTCWT 辞書の相互コヒーレンス (mutual coherence)⁽⁶³⁾ などを考慮する必要がある、これも今後の課題である。

4.3.4 Noise Shaping 法における新しい係数選択手法の提案

第4.3.1節の Noise Shaping 法では、式(4.1)の閾値処理において複素係数単位で係数を選択していた。しかし、式(4.1)に対応する l^0 ベース交互射影法を用いた疎表現近似解法における $ht(\alpha, K)$ は、最大値より K 個の非ゼロ係数を選択する操作であり、必ずしも複素係数単位で係数を選択する必要はない。そこで、式(4.1)において複素数単位で選択するという制約を排除し、式(4.9)のように実部・虚部からそれぞれ独立に係数を選択する手法を提案する⁽⁴⁶⁾⁽⁶⁴⁾。

$$\hat{y}_r = \begin{cases} 0 & \text{if } |y_r| < \theta \\ y_r & \text{otherwise} \end{cases} \quad \hat{y}_i = \begin{cases} 0 & \text{if } |y_i| < \theta \\ y_i & \text{otherwise} \end{cases} \quad (4.9)$$

疎表現近似解法における提案手法の有効性は次の2通りの解釈により示すことができる。1つは、係数選択の際の誤基底の排除である。第4.3.3節の考察で述べたように、Noise Shaping 法のアプローチは貪欲アルゴリズムである。このため、近似精度の高い基底から優先的に選択されていく。また、1度選択した基底は収束時まで保持される。従来手法では、実部と虚

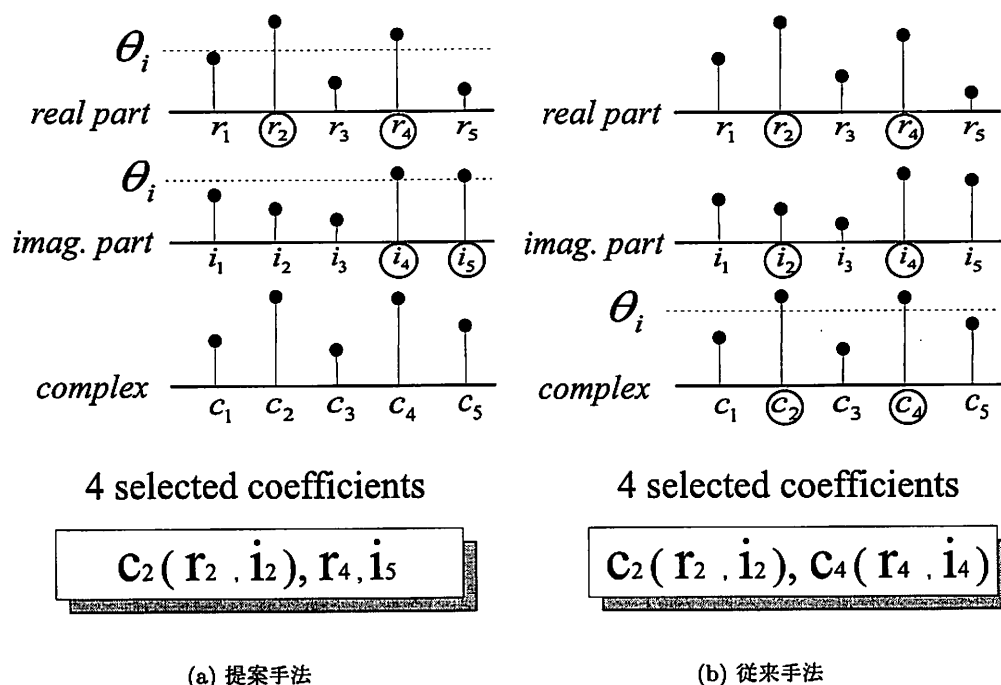


図 4-5: 提案手法と従来手法の係数選択の違い

部の各基底をまとめて複素数の基底として扱っている。そのため、実部と虚部のどちらか一方のみが優先順位の高い基底であった場合も、2つの基底を選択してしまう。これに対して、提案手法では、実部と虚部の基底を独立に考えることで、優先順位の高い基底のみが選択される。図4-5は従来手法と提案手法の係数選択の違いをあらわしている。図4-5において、 r_k は実部の基底、 i_k は虚部の基底、 c_k は複素数の基底である。図4-5(b)の従来手法では、複素数基底の係数に対して閾値処理をするため、この例では $c_2(r_2, i_2)$ 、 $c_4(r_4, i_4)$ の4つの基底が選択される。このとき i_2 はそれほど優先順位が高くないにもかかわらず選択されてしまっている。このように、実部基底と虚部基底の一方のみの優先順位が高い場合でも、両基底を選択してしまう可能性がある。これに対して、図4-5(a)の提案手法では、実部基底と虚部基底を独立と扱うことで、 i_2 よりも優先順位の高い i_5 が選択されている。

もう1つは、辞書 $\mathbf{D}$ の定義である。疎表現の性能は近似解法の性能に加えて、辞書の定義により決まる。これは信号を表現するのに適切な基底集合が辞書 $\mathbf{D}$ として定義されていなければ、その疎表現の性能も悪くなるためである。提案手法と従来手法は辞書 $\mathbf{D}$ の違いとして解釈することができる。第3.6.2節で述べたようにDTCWTは2次元では方向基底の集合である。この点では従来手法と提案手法は同じであり、2次元信号に適した基底集合である。ここで、方向基底の複素ウェーブレットを $\psi^c = \psi^r + j\psi^i$ とすると、従来手法では辞書は $\mathcal{D}_c = \{\psi_i^c \in \mathbb{R}^n, i = 1, \dots, 2n\}$ として定義される。一方で、提案手法では

$\mathcal{D}_p = \{\psi_i^r, \psi_i^i \in \mathbb{R}^n, i = 1, \dots, 2n\}$ となる。ここで、 ψ^r と ψ^i の間にはヒルベルト変換の関係があるため、これらの基底はシフト関係にある。このため、提案手法ではより密な方向基底の集合を辞書 $\mathcal{D}_p$ として定義していることになる。基底集合が密になるほど信号をより高い精度で近似する基底が増えるため、提案手法では疎表現の性能が向上する。

4.4 疎表現係数列を用いた静止画像符号化実験

4.4.1 Noise Shaping 法のパラメータと疎表現の性能

Noise Shaping 法における 1 反復当たりの非ゼロ係数の増加個数を表現するために 2 つのパラメータを定義する。1 つは反復回数を制御するパラメータ *Iteration*、もう 1 つは終了時の非ゼロ個数を制御するパラメータ *NonZero* である。*NonZero* は式 (3-41) における λ に相当する。これら 2 つのパラメータより Noise Shaping 法における 1 反復当たりの非ゼロ係数の増加個数 *GetCoeff* は式 (4-10) となる。

$$GetCoeff = \frac{NonZero}{Iteration} \quad (4-10)$$

各パラメータと求まる疎表現の特徴を以下に示す。

パラメータ *Iteration* と疎表現

反復終了時の非ゼロ個数を *NonZero* = 3000 と固定し、反復回数 *Iteration* を変動させた場合の非ゼロ係数へのエネルギー集約度の結果を図 4-6 に示す。また、*Iteration* = 90 における疎表現係数列の絶対値の相互情報量を表 4-1、表 4-2 に示す。図 4-6 より、反復回数が多いほど非ゼロ係数へのエネルギー集約度高くなる。これより良好な疎表現の近似解を得るためには反復回数を増やす必要がある。実験的に 40 回程度の反復からは大きな違いは見られない。式 (4-10) より、反復回数が多いほど各反復において選択する非ゼロ係数の個数 *GetCoeff* が小さくなる。このため誤った基底を選択する確率が小さくなり、このような結果となる。また、表 3-2、表 3-4 と表 4-1、表 4-2 を比較すると疎表現係数列において相互情報量が小さくなっていることがわかる。これより、疎表現を求めることにより係数間の冗長性が削減される。すなわち、疎表現と係数間の冗長性にはトレードオフが存在する。

パラメータ *NonZero* と疎表現

反復回数 *Iteration* = 50 と固定し、終了時の非ゼロ個数 *NonZero* を変動させた場合の非ゼロ係数へのエネルギー集約度を図 4-7 に示す。図 4-7 より、曲線は *NonZero* までは高い傾き、*NonZero* 後では低い傾きで変化している。これは *NonZero* までの非ゼロ係数で疎表現の近似解を生成するため、それ以降の係数列は疎にならないためである。

表 4-1: Lena における疎表現係数列の絶対値の相互情報量 (文献⁽⁴⁶⁾ 参照)

subband	Absolute of DTCWT(Iteration=90 ,nz=3000)								
	Level1			Level2			Leve3		
	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)
15	0.106	0.246	0.406	0.174	0.346	0.560	0.066	0.215	0.501
45	0.074	0.188	0.253	0.193	0.298	0.570	0.093	0.262	0.611
75	0.165	0.418	0.474	0.100	0.467	0.643	0.045	0.227	0.372
-15	0.107	0.239	0.319	0.188	0.366	0.648	0.068	0.251	0.569
-45	0.102	0.208	0.299	0.201	0.337	0.663	0.119	0.309	0.689
-75	0.152	0.383	0.461	0.109	0.479	0.639	0.041	0.242	0.377

表 4-2: Barbara における疎表現係数列の絶対値の相互情報量

subband	Absolute of DTCWT(Iteration 90 ,nz=3000)								
	Level1			Level2			Leve3		
	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)	I(X;PX)	I(X;CX)	I(X;NX)
15	0.109	0.273	0.646	0.114	0.278	0.734	0.087	0.136	0.521
45	0.129	0.371	0.649	0.118	0.280	0.989	0.089	0.140	0.630
75	0.246	0.677	1.007	0.101	0.481	0.887	0.042	0.225	0.479
-15	0.097	0.215	0.642	0.139	0.254	0.752	0.076	0.158	0.550
-45	0.183	0.413	0.686	0.115	0.314	0.878	0.122	0.193	0.682
-75	0.200	0.668	0.943	0.131	0.410	0.850	0.060	0.214	0.499

4.4.2 非ゼロ係数へのエネルギー集約度による提案手法の性能

第 4.3.4 節において提案した係数選択手法と従来の係数選択手法の非ゼロ係数へのエネルギー集約度の違いを図 4-8 に示す。参考までに、図 4-8 には疎表現前の DTCWT 係数と DWT 係数の非ゼロ係数へのエネルギー集約度も示している。提案手法は、Lena において従来手法や DWT よりも約 1.5[dB]-2.0[dB]、Barbara において従来手法より約 1.0[dB]-2.0[dB]、DWT より約 1.0[dB]-3.0[dB] 程度の高いエネルギー集約度が得られる。また、疎表現後の DTCWT 係数列は疎表現前の DTCWT 係数列と比較し、最大で 7.0[dB] 程度のエネルギー集約度の改善となる。これより、提案手法は非ゼロ係数へのエネルギー集約度という観点において、従来手法や DWT より優位である。

4.4.3 全符号量による提案手法の性能

第 4.4.2 節より、DTCWT の疎表現係数列は DWT よりも高いエネルギー集約度となることがわかった。しかし、2D DTCWT は 2D DWT と比較して、4 倍の係数が発生している。そこで、これらの係数の個数も考慮した全符号量における評価を行う。 $N \times N$ サンプルの入力画像に対して、変換後の係数が $M \times M$ になるとする。2D DWT では $M = N$ 、2D DTCWT では $M = 2N$ である。まず、レベル i におけるエントロピー H_i を算出する。エントロピーは式 (4-11) により計算される。

$$H_i = - \sum_x P_i(x) \log_2 P_i(x) \quad (4-11)$$

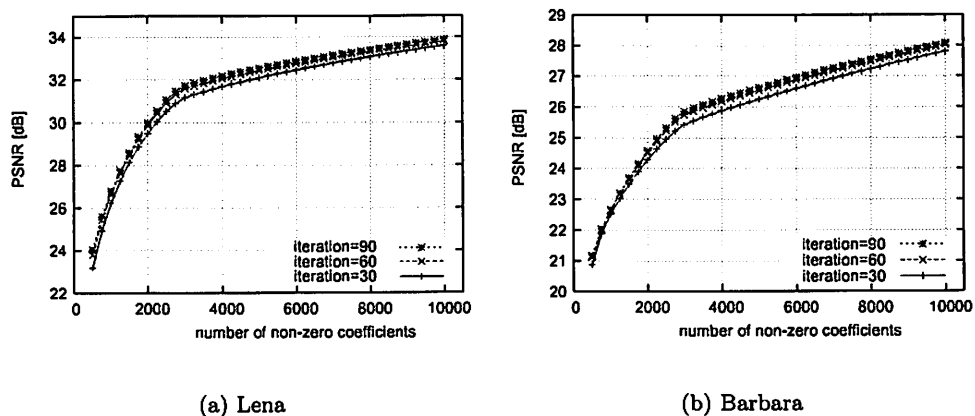


図 4-6: パラメータ *Iteration* と非ゼロ係数へのエネルギー集約度 (文献⁽⁴⁶⁾ 参照)

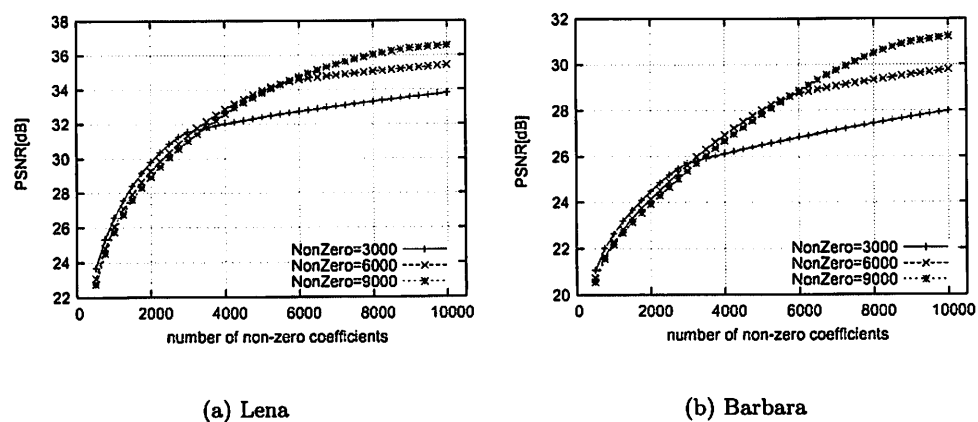


図 4-7: パラメータ *NonZero* と非ゼロ係数へのエネルギー集約度 (文献⁽⁴⁶⁾ 参照)

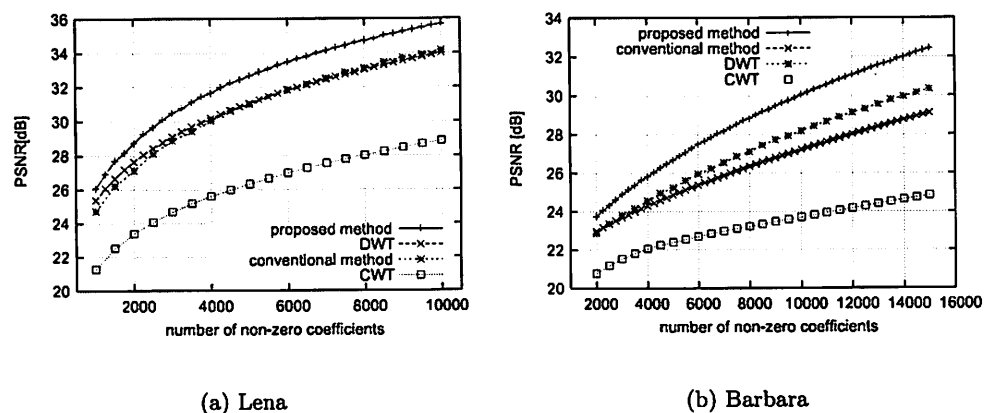


図 4-8: 提案手法と従来手法の非ゼロ係数へのエネルギー集約度 (文献⁽⁴⁶⁾ 参照)

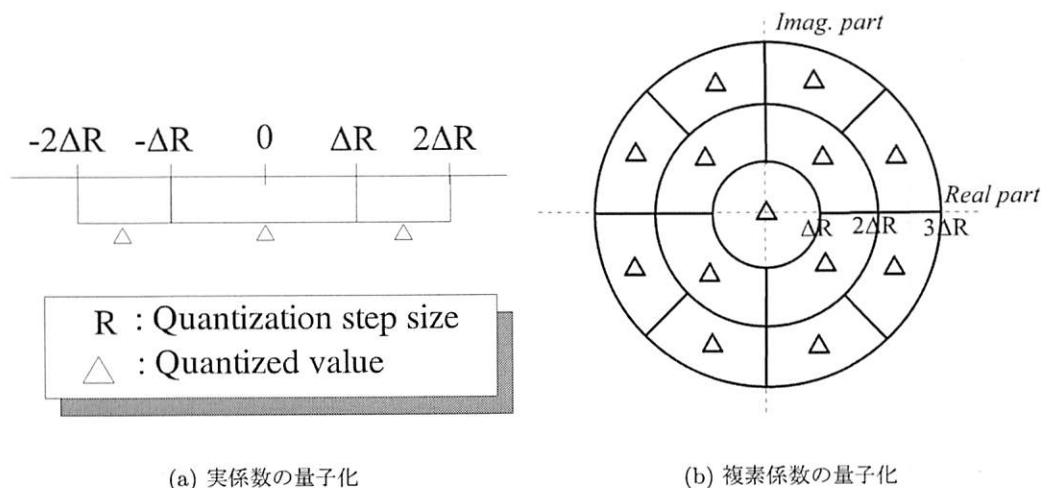


図 4-9: 実係数と複素係数の量子化

ここで、 $P_i(x)$ はレベル i における係数の発生確率である。log の底が 2 の場合は、エントロピーは可変長符号化を行ったときの平均符号長の下限となり、1 係数を符号化するのに必要なビット数を示している。よって、式 (4.11) はレベル i に属する 1 係数を符号化するのに必要なビット数である。エントロピーに係数の個数を乗算することで全符号量を求めることができる。これより、全符号量は各レベルにおけるエントロピーの和より式 (4.12) となる。

$$R = \frac{(\sum_{i=1}^{L} H_i) \cdot M^2}{N^2} \quad (4.12)$$

ここで、 L はレベル数、 R は 1 画素あたりの符号量をあらわしている。式 (4.11) における係数の発生確率 $P_i(x)$ を求めるためには、係数の量子化処理を行う必要がある。複素係数の量子化処理は 2 種類考えることができる。1 つは図 4-9(a) のように実係数に対してスカラー量子化を用いる場合である。この場合、複素係数の実部と虚部を独立に符号化することになる。もう 1 つは図 4-9(b) のように複素係数に対してベクトル量子化を用いた場合である。この場合、複素空間上で定義される同心円を区分することで量子化される。図 4-9 における量子化ステップ ΔR を変更することで、全符号量と歪みのトレードオフが実現される。

図 4-10(a)、図 4-11(a)、図 4-12(a) はそれぞれ Lena (512×512: グレースケール)、Barbara (512×512: グレースケール)、Goldhill (512×512: グレースケール) における全符号量のレート歪み特性である。従来手法は疎表現の近似解を求める際に、複素数として係数が選択されているため、複素数のスカラー量子化とベクトル量子化の両方を適用した。一方で、提案手法は実部と虚部を独立に選択しているためスカラー量子化のみを適用した。また、DWT の SPIHT 符号化の結果を同時に示す。図 4-10(a)、図 4-11(a)、図 4-12(a) より従来手法のスカラー量子化が最も性能が悪く、従来手法のベクトル量子化が最も性能が良い。これより、DTCWT の疎表現係数列では適切なベクトル量子化を用いることで符号化効率の改善が期待できる。従来手法のベクトル量子化は SPIHT 符号化と比較し、Lena では低ビットレート

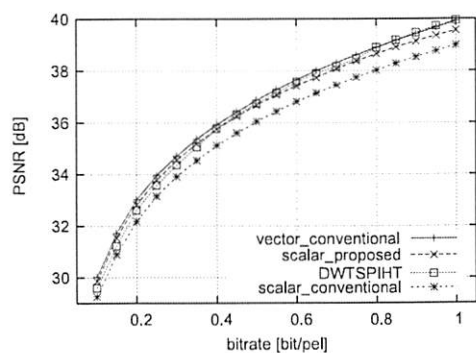
において PSNR が約 0.5 [dB] 高い。Barbara ではすべてのビットレートにおいて PSNR が約 0.5[dB]–1.0[dB] 高い。Goldhill では、PSNR はほぼ同じである。また、提案手法のスカラ量子化は、従来手法のベクトル量子化よりは性能が低い。Lena, Barbara では低ビットレートにおいて SPIHT 符号化より約 0.2[dB]–0.5[dB] PSNR が高く、Goldhill は SPIHT 符号化とほぼ同等の PSNR である。

図 4-10, 図 4-11, 図 4-12 の (b)–(f) はそれぞれ各画像の原画像、特定ビットレートにおける DTCWT の再構成画像、DWT の再構成画像、DTCWT の局部拡大画像、DWT の局部拡大画像である。ビットレートは、Lena は 0.3[bit/pel], Barbara は 0.4[bit/pel], Goldhill は 0.4[bit/pel] である。Lena において、DTCWT は鏡の縁や帽子の縁など斜め方向のエッジを高い近似精度で滑らかに再構成し、エッジの輪郭に沿った雑音が発生している。一方で、DWT はチェッカーボードアーティファクトの影響よりエッジ近辺でジャギーのような雑音が発生している。また、帽子のテクスチャに関して DTCWT の方が高い精度で復元しており優位である。しかし、垂直方向のエッジは DWT の方が精度良く再構成している。これは DWT では垂直方向のエッジを分離する HL サブバンドが存在するが、DTCWT は垂直方向のエッジを±75°方向の2つのサブバンドへ分離してしまうためである。水平方向についても同様である。また、Barbara においても、本の輪郭などは DTCWT の方が精度良く再構成されているが、机の脚のような垂直方向のエッジは DWT の方が高い精度で再構成している。特に、ネクタイのような縞模様は DTCWT の優位性が顕著であり、非常に高い精度で再構成している。一方で、DWT ではチェッカーボードアーティファクトの影響より格子状の模様として表現されており、本来の構造と異なっている。Goldhill についても同様の考察ができる。

4.4.4 疎表現を用いた静止画像符号化の考察

以上の実験結果に基づき DTCWT の疎表現を符号化へ適用する際の考察を行う。まず、第 4.4.1 節において述べた 2 つのパラメータの決定手法について述べる。反復回数を制御するパラメータである *Iteration* は、疎表現の近似解の性能および係数間の冗長性に影響を与えていた。反復回数はエンコードにおける処理時間の大部分を占めるため、処理時間を考慮して決定する必要がある。反復回数が少ない場合、疎表現の近似解の精度が悪いため非ゼロ係数へのエネルギー集約度は低いが、係数間に高い冗長性が残るため、冗長性を削減するボスト処理が効果的に機能すると予想される。

次に、反復終了時の非ゼロ個数を制御するパラメータである *NonZero* と疎表現の関係を述べる。ある非ゼロ個数に最も多くのエネルギーを集約するためには、その個数を *NonZero* として疎表現係数列を求める必要がある。つまり、ある疎表現係数列が最大のエネルギー集約度を示すのは一部の非ゼロ個数のみであり、それ以外の非ゼロ個数における最大のエネルギー集約度は他の疎表現係数列から与えられる。これは信号の表現方法が 1 通りしか存在しない DWT と大きく異なる。このような疎表現の特性より、DTCWT の疎表現における静止画像符号化は SNR スケーラブル符号化よりも特定の符号量において最適な符号化を



(a) レート歪み特性



(b) 原画像



(c) DTCWT 34.71 (dB)



(d) DWT 34.36 (dB)

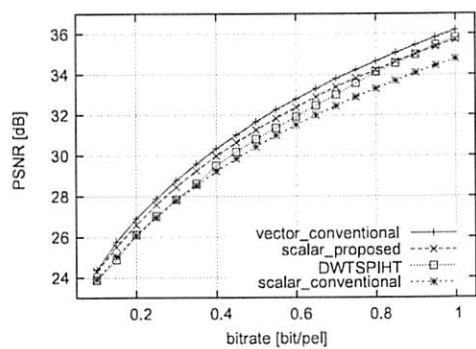


(e) DTCWT の局部拡大画像

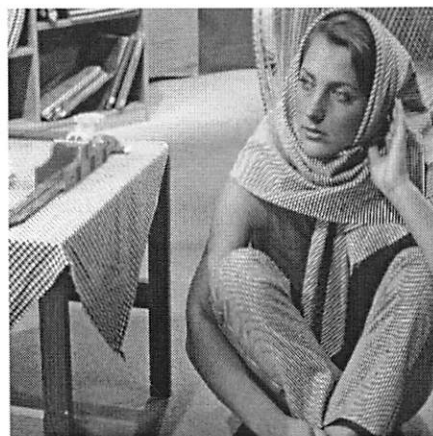


(f) DWT の局部拡大画像

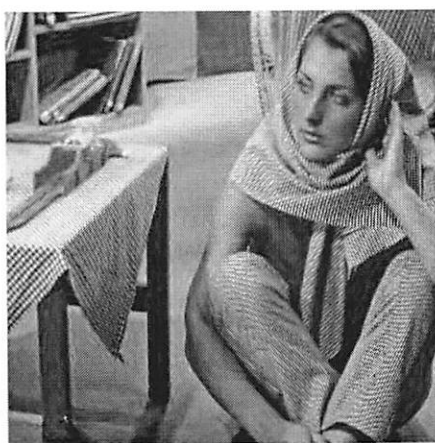
図 4-10: Lena における並列木複素ウェーブレット疎表現係数列の符号化特性



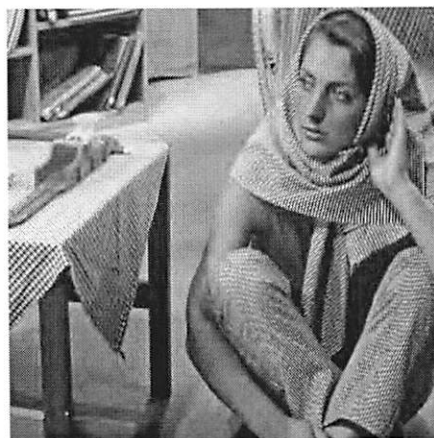
(a) レート歪み特性



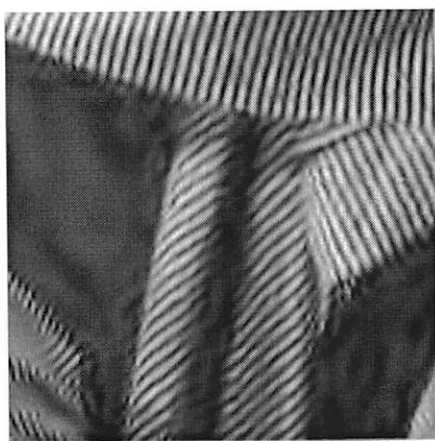
(b) 原画像



(c) DTCWT 30.34 (dB)



(d) DWT 29.51 (dB)

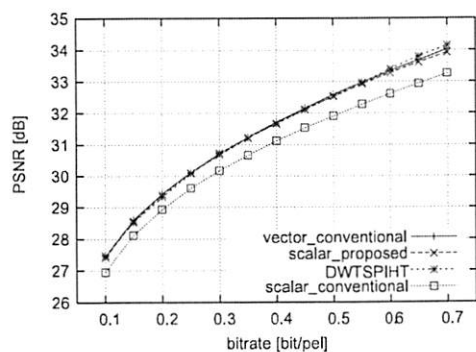


(e) DTCWT の局部拡大画像



(f) DWT の局部拡大画像

図 4-11: Barbara における並列木複素ウェーブレット疎表現係数列の符号化特性



(a) レート歪み特性



(b) 原画像



(c) DTCWT 31.72 (dB)



(d) DWT 31.65 (dB)



(e) DTCWT の局部拡大画像



(f) DWT の局部拡大画像

図 4-12: Goldhill における並列木複素ウェーブレット疎表現係数列の符号化特性

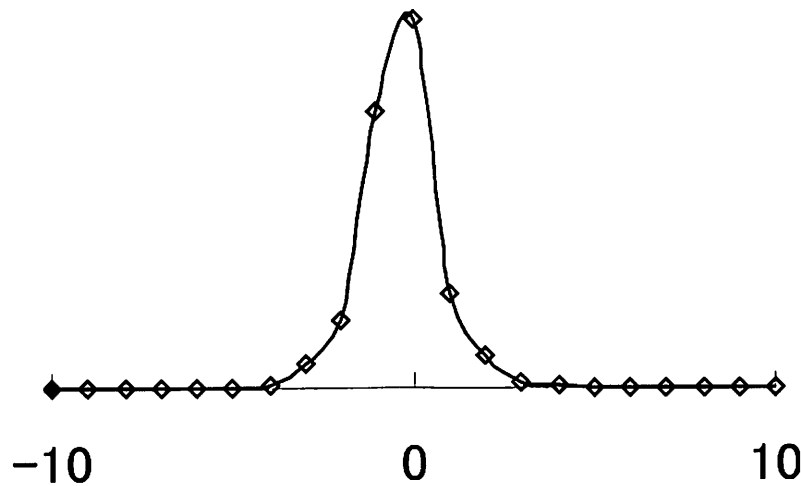


図 4-13: ステップエッジに対する並列木複素ウェーブレット係数の絶対値応答

行うことが望ましいと考えられる。ただし、疎表現係数列において SNR スケーラブル符号化を実現すること自体は可能である。そして、この特定の符号量より逆算してパラメータ *Nonzero* を決定し、そのパラメータに基づき Noise Shaping 法を適用し、疎表現係数列を求める。

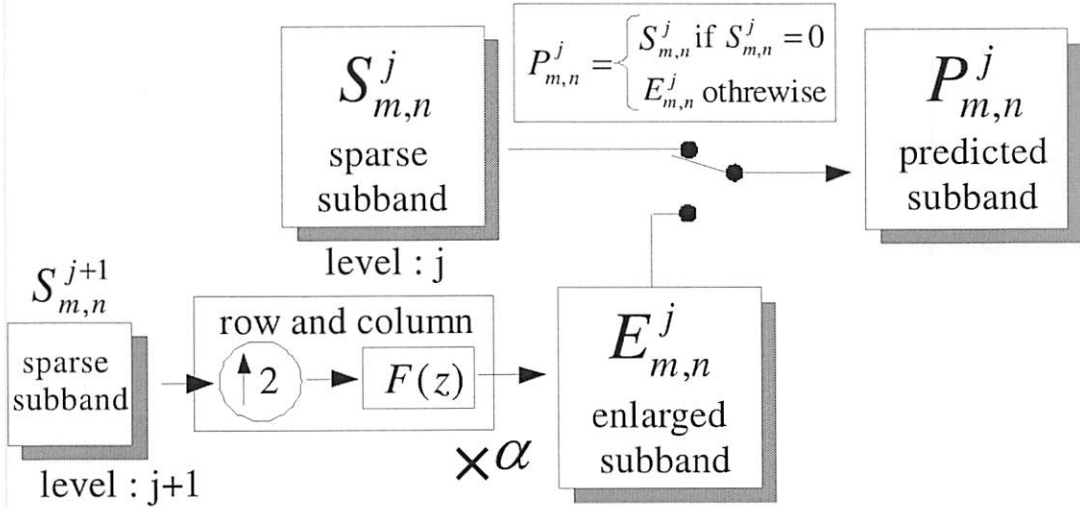
次に、疎表現と量子化について考察を行う。第 4.4.3 節の実験結果より、スカラ量子化を用いた場合は非ゼロ係数への高いエネルギー集約度が得られる提案手法が有効であった。しかし、複素係数のベクトル量子化を用いた場合は従来手法が最も高い符号化効率となった。これは、従来手法における非ゼロ係数へのエネルギー集約の悪さを克服するほどの効果的な係数のベクトル量子化が行われたからである。これより、静止画像符号化において疎表現の近似解を求めるときは、エネルギー集約と優位係数の存在位置を考慮しなければならない。よって、提案手法と従来手法を統合するなどして、ある程度の良好なエネルギー集約とベクトル量子化が効果的に機能する係数列を生成することが今後の課題となる。

4.5 疎表現係数列における係数予測手法

4.5.1 疎表現係数の絶対値予測手法

第 3.7.1 節において、疎表現前の係数列の絶対値には多くの冗長性が存在することを述べた。また、第 4.4.1 節において、疎表現係数列の冗長性は疎表現前より削減されることを述べた。しかし、表 4-1 と表 4-2 より、疎表現係数列の絶対値には依然として冗長性が残っている。そこで、この冗長性を利用する手法を検討する。本提案手法は復号化器において受信した少数の非ゼロ係数から非受信係数の絶対値を予測する手法である。

ステップエッジに対するレベル 3 の DTCWT 係数の絶対値を図 4-13 に示す。図 4-13 よ


 図 4-14: 疎表現係数列における係数予測手法 (文献⁽⁶⁶⁾ 参照)

り, DTCWT 係数の絶対値はガウス関数で近似される. また, 低解像度レベルの絶対値ほど分散の大きなガウス関数で近似される特性がある. これより, DTCWT 係数の絶対値はガウシアンピラミッド構造で近似することができる. ガウシアンピラミッドは, 原信号にガウシアンフィルタ (W) 処理とダウンサンプリング処理 (DS) を 1 組とする処理を 1 回, 2 回, $\dots, N$ 回 と繰り返すことで得られたガウシアン成分 $G_0, G_1, \dots, G_N$ の連なりである. つまり, ガウシアン成分 G_{n+1} は次式により与えられる.

$$\bar{G}_{n+1} = W * G_n \quad (4-13)$$

$$G_{n+1} = DS(\bar{G}_{n+1}) \quad (4-14)$$

ここで $*$ は畳み込み演算である. このときに失われる高周波成分の連なりはラプラシアンピラミッドと呼ばれる⁽⁶⁵⁾. ガウシアンピラミッドは低周波数成分の連なりから構成されるため, 低解像度ピラミッドの単純な拡大補間処理により, 高解像度ピラミッドを精度良く予測できる. この特性を利用し, 低解像度サブバンドの拡大補間処理により高解像度のサブバンドを予測する.

サブバンドの予測手法を図 4-14 に示す. 図 4-14 の処理を各レベル, 各方向サブバンドについて低解像度から順に独立に行う. まず, 低域サブバンドの係数 $S_{m,n}^{j+1}$ を零補間とローパスフィルタを用いて拡大する. j は解像度レベルであり, (m, n) は係数の座標である. 次に, 同一のレベル・方向で固定値の重み係数 α を用いてエネルギーを変更し, 拡大サブバンドの係数 $E_{m,n}^j$ を得る. レベル j のサブバンドの係数 $S_{m,n}^j$ が非受信係数 (ゼロ係数) の場合, 拡大サブバンドの係数 $E_{m,n}^j$ で置き換え, 予測サブバンドの係数 $P_{m,n}^j$ を生成する.

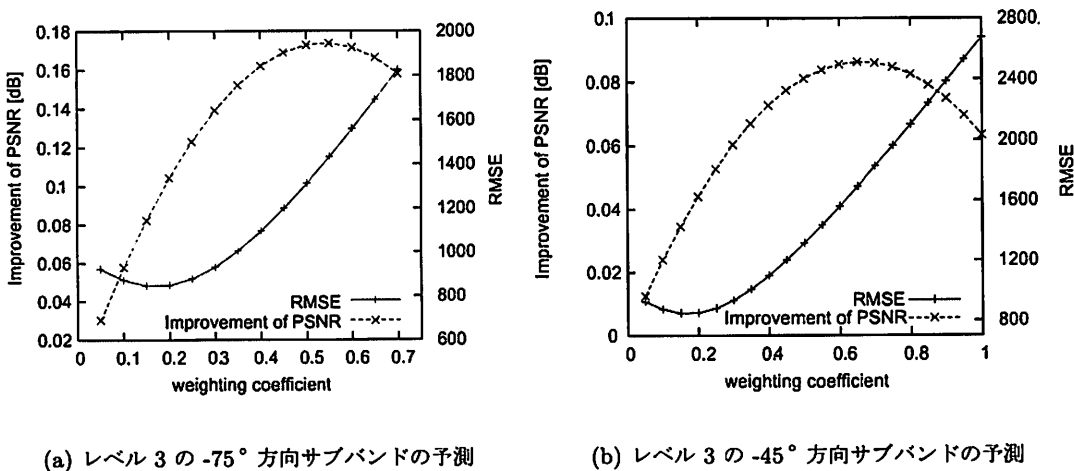


図 4-15: 係数予測における重み係数と RMSE・PSNR 向上値 (文献⁽⁶⁴⁾ 参照)

表 4-3: Lena における係数予測手法の PSNR 達成率 (文献⁽⁶⁴⁾ 参照)

predicted level	direction of subband						All
	15°	45°	75°	-15°	-45°	-75°	
from 5 to 4 (%)	32.9	27.3	48.6	37.1	39.4	40.2	34.7
from 4 to 3 (%)	50.0	50.0	57.1	56.7	59.2	54.4	47.8

4.5.2 予測手法の評価実験

重み係数 α を変化させたときの、DTCWT 係数の真値と予測値の RMSE と予測値からの再構成画像の PSNR 向上値を図 4-15 に示す。冗長変換である DTCWT は直交変換と異なり変換係数間に依存性がある。このため、DTCWT 係数において最小誤差を与える予測値が、再構成画像で最大の PSNR を与えないという特性がある。これは、DTCWT 係数に適した量子化を考える上で非常に重要な特性であり、興味深い特性といえる。

また、適切な重み係数 α を用いてレベル 4 とレベル 3 のサブバンドを予測した結果を表 4-3 と表 4-4 に示す。表 4-3 と表 4-4 は、真値を用いたときの PSNR 向上に対する提案手法による PSNR 向上の割合であり、達成率である。本研究では、まだ位相の予測手法が確

表 4-4: Barbara における係数予測手法の PSNR 達成率

predicted level	direction of subband						All
	15°	45°	75°	-15°	-45°	-75°	
from 5 to 4 (%)	33.6	42.3	30.2	33.0	39.8	36.1	32.9
from 4 to 3 (%)	30.3	21.8	36.3	32.3	25.3	37.4	31.5

立していないため、予測した絶対値の位相には最低限の情報として、正負の符号情報に相当する 2[bit] を与えている。表 4-3 と表 4-4 より、本予測手法により 35～50[%] 程度の達成率が得られており、復号化器における疎表現の絶対値の予測が可能であることがわかる。

4.6 むすび

本章では、DTCWT の疎表現係数列を用いた静止画像符号化の検討を行った。まず、DTCWT の疎表現係数列を用いた符号化システムを述べた。次に、Noise Shaping 法の説明をし、これが疎表現の近似解を求める l^0 ノルムベース交互射影法と等価であることを示した。そして、等価性より Noise Shaping 法における新しい係数選択手法を提案した。実験より、提案手法の性能と DTCWT の疎表現係数列の符号化性能を確認した。最後に、復号化器における DTCWT の疎表現係数列の予測手法を提案し、その性能を評価した。

第5章

結論

5.1 総括

本研究では、DTCWT の疎表現係数列を用いた静止画像符号化の検討を行った。国際標準画像符号化の変換方式である DCT や DWT は直交変換である。これに対して、DTCWT はシフト不変性が成立し、方向分離特性をもつ冗長変換である。本研究では、このような変換特性をもつ DTCWT の疎表現に基づく静止画像符号化を実現し、直交変換による静止画像符号化よりも高い符号化効率を得ることを目的とした。

本研究では、まず、DTCWT における雑音変形手法である Noise Shaping 法が、疎表現の近似解を求める l^0 ノルムベース交互射影法と等価であることを示した。そして、その等価性より Noise Shaping 法において新しい係数選択手法を提案した。実験より、提案手法は非ゼロ係数へより多くのエネルギーを集約することを確認し、従来手法よりも正確な疎表現の近似解を求めていることを示した。また、DTCWT の疎表現係数列を用いた静止画像符号化は、全符号量において DWT の SPIHT 符号化よりも高い符号化効率となることを確認した。そして、DTCWT 画像符号化の符号化効率改善の一手法として、疎表現係数列の絶対値に残る冗長性を利用した係数予測手法を提案した。実験より、本予測手法により PSNR の向上が可能であることを示した。

第1章では、本研究の背景と目的、および本論文の構成について述べた。

第2章では、静止画像の変換符号化の現状について述べた。まず、国際標準画像符号化は直交変換を用いた変換符号化であることを述べ、JPEG と JPEG 2000 の概要を述べた。次に、JPEG 2000 の変換方式である DWT について、フィルタバンクによる実装方法や完全再構成条件について述べた。そして、DWT の問題点を示した。最後に、DWT の SPIHT 符号化方式について述べ、その符号化効率を示した。

第3章では、DTCWT による新しい静止画像符号化について述べた。まず、2種類の CWT について述べ、ヒルベルト変換を説明した。次に、DTCWT のフィルバンクの構造やフィルタの設計方法について述べた。そして、DTCWT の性質と冗長性を示した。さらに、この係数列に対する SPIHT 符号化は高い符号化効率とならないことを確認し、疎表現の必要性を述べた。最後に、疎表現の概念や理論を述べ、疎表現の近似解を求める手法を述べた。

第4章では、DTCWT の疎表現係数列を用いた静止画像符号化の検討を行った。まず、

DTCWT の疎表現係数列を用いた符号化システムの概要を述べた。次に、Noise Shaping 法の説明をし、これが疎表現の近似解を求める l^0 ノルムベース交互射影法と等価であることを示した。そして、等価性より Noise Shaping 法における新しい係数選択手法を提案した。実験より、提案手法の性能と DTCWT の疎表現係数列の符号化性能を確認した。最後に、復号化器における係数予測手法を提案し、その性能を評価した。

第5章は結論であり、本論文の総括および今後の課題について述べている。

5.2 今後の課題

今後の課題として以下がある。

5.2.1 非冗長複素ウェーブレット変換の検討

本研究では、冗長複素ウェーブレット変換である DTCWT の疎表現係数列を用いた画像符号化の実現を目的とした。一方で、第3.2節で述べた非冗長複素ウェーブレット変換である MPCWT を用いた画像符号化手法も興味深い。MPCWT は非冗長変換であるため、DWT と同様に変換後のデータ数は不変であるが、 $\pm 15^\circ$ 、 $\pm 45^\circ$ 、 $\pm 75^\circ$ の6方向を異なるサブバンドに分離する変換である。このため、MPCWT は DWT と比較し、高いエネルギー集約度が得られることが予想される。また、疎表現のような理論が必要ないため、DWT を用いた画像符号化の理論が適用できる可能性が高い。

5.2.2 疎表現の近似解法の検討

疎表現問題は NP 困難な問題である。そのため、この問題を現実的な時間で解くためには近似手法が必要である。本研究では、疎表現近似手法として l^0 ノルムベース交互射影法を用いた。しかし、この他の近似手法として BP や MP などがあり、DTCWT の辞書におけるこれらの手法の評価が必要である。また、疎表現問題は最適化問題の一種である。この観点から、さらに高速で近似解の精度の高いアルゴリズムの追求が必要である。

5.2.3 疎表現係数列のベクトル量子化とコンテキストモデリングの検討

第4.3.3節では、疎表現係数列にはスカラ量子化よりベクトル量子化のほうが望ましいことを述べた。そこで、疎表現係数列の特性をさらに追及し、適切なベクトル量子化の検討が必要である。また、疎表現係数列の位置情報を効率的に符号化するコンテキストモデリングの検討も必要である。

5.2.4 疎表現係数列における位相予測手法の検討

DTCWT の疎表現係数列における絶対値の予測に関しては、第 4.5 節で提案した手法により一定の精度で予測可能である。しかし、DTCWT の疎表現係数列の位相の予測については未検討であり、第 3.7.2 節の位相の冗長性を基に、疎表現係数列における位相の冗長性を調査し、位相の予測手法を検討する必要がある。

謝辞

本研究の機会及び素晴らしい実験環境を与えて下さり、貴重な時間を割いて研究の方向性を御指導頂きました渡辺 裕 教授に心から感謝致します。

また、学部時代の指導教授であり、研究の方向性について、数々の有意義な御意見、御助言を賜りました早稲田大学国際情報通信研究科富永英義教授に深く感謝致します。

本研究を進めるにあたり、日頃から惜しみなく御指導して頂き、文章の書き方等でお世話になりました河村 圭氏に心から感謝いたします。

信号処理の理論などをご教授いただき、常に興味深いご意見をいただいた石川 孝明氏に心から感謝いたします。

貴重な御意見、様々な御提案を頂いた渡辺研究室の皆様に御礼申し上げます。特に、学部時代からの同級生とともに励ましあった東 昌史氏とともに困難を乗り越えてきた石井 大祐氏に心から感謝いたします。

学部時代からお世話になった富永研究室の皆様に、御礼申し上げます。特に、3年間公私にわたり、共に頑張ってきた同級生のみなに感謝致します。

最後に、私をここまで育てて下さった家族に深く感謝します。

平成 20 年 2 月 5 日

参考文献

- (1) “パソコンテレビ GyaO [ギャオ],” <http://www.gyao.jp/>
- (2) “flickr,” <http://www.flickr.com/>
- (3) “YouTube,” <http://jp.youtube.com/>
- (4) “ニコニコ動画 (RC2),” <http://www.nicovideo.jp/>
- (5) “ソニー製品情報,” <http://www.sony.jp/>
- (6) “NHK 放送技術研究所,” <http://www.nhk.or.jp/strl/>
- (7) “PlayStation.com (Japan),” <http://www.jp.playstation.com/index.html>
- (8) “任天堂ホームページ,” <http://www.nintendo.co.jp/index.html>
- (9) 安田 浩, 渡辺 裕, “ディジタル画像圧縮の基礎,” 日経 BP 出版センター, Jan. 1996.
- (10) 福原 隆浩, 板倉 英三郎, “JPEG 2000 詳細解説,” CQ 出版, Aug. 2004.
- (11) “HD Photo,” <http://www.microsoft.com/windows/windowsmedia/forpros/wmphoto/>
- (12) 小野 定康, 鈴木 純司, “わかりやすい JPEG 2000 の技術,” オーム社, May 2003.
- (13) 貴家 仁志, “JPEG 2000 を中心とした画像圧縮技術の新しい流れ,” *Interface*, pp. 46–58, Jan. 2002.
- (14) 貴家 仁志, 渡辺 修, “JPEG 2000 符号化アルゴリズムの要素技術,” *Interface*, pp. 59–71, Jan. 2002.
- (15) 榊原 進, “ウェーブレットビギナーズガイド”, 東京電機大学出版局, 1995.
- (16) 新井康平, “ウェーブレット解析の基礎理論”, 森北出版, 2000.
- (17) 中野, 山本, 吉田, “ウェーブレットによる信号処理と画像処理”, 共立出版, 2000.
- (18) 貴家 仁志, “マルチレート信号処理”, 昭晃堂, 1995.
- (19) N. G. Kingsbury, “Complex wavelets for shift invariant analysis and filtering of signals,” *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, May 2001.
- (20) N. G. Kingsbury, and J. F. A. Magarey, “Wavelet transform in image processing,” *Proceedings 1st European Conference Signal Analysis and Prediction*, pp. 23–34, June, 1997.

- (21) J. M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Transactions Signal Processing*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.
- (22) A. Said, and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–249, June 1996.
- (23) S. Joo, H. Kikuchi, S. Sasaki, and J. Shin, "A new image coding technique with low entropy using a flexible zerotree," *IEICE Transactions on Communications*, vol. E81–B, no. 12, pp. 2528–2535, Dec. 1998.
- (24) J. Romberg, M. Wakin, H. Choi, and R. G. Baraniuk, "A geometric hidden markov tree wavelet model," *Proceedings SPIE Wavelets : Applications in Signal and Image Processing X*, pp. 80–86, San Diego, 2003.
- (25) J. K. Romberg, H. Choi, and R. G. Baraniuk, "Multiscale edge grammars for complex wavelet transforms," *Proceedings IEEE International Conference Image Processing*, vol. 1, pp. 614–617, Thessaloniki, Oct. 2001.
- (26) N. G. Kingsbury, "Image processing with complex wavelets," *Philosophy Transaction Royal Society London A, Math. Phys. Sci.*, vol. 357, no. 1760, pp. 2543–2560, Sept. 1999.
- (27) N. G. Kingsbury, "The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters," *Proceedings 8th IEEE DSP Workshop*, no. 86, Utah, Aug. 9–12, 1998.
- (28) I. W. Selesnick, "Hilbert transform pairs of wavelet bases," *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 170–173, June 2001.
- (29) H. Kawabata, H. Toda, Z. Zhang, and H. Fujiwara, "A new complex wavelet transform by using RI-spline wavelet," *Proceedings IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 937–940, Montreal, May 2004.
- (30) F. Fernandes, R. Spaendonck, and C. Burrus, "A new framework for complex wavelet transforms," *IEEE Transactions Signal Processing*, vol. 51, no. 7, pp. 1825–1837, July, 2003.
- (31) F. Fernandes, R. Spaendonck, and C. Burrus, "Multidimensional, mapping-based complex wavelet transforms," *IEEE Transactions Image Processing*, vol. 14. no. 1, Jan. 2005.

- (32) 戸田, 章, 川畑, “最新ウェーブレット実践講座 入門と応用”, ソフトバンククリエイティブ, 2005.
- (33) I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, “The dual-tree complex wavelet transform,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123-151, Nov. 2005.
- (34) H. Ozkaramanli, and R. Yu, “On the phase condition and its solution for hilbert transform pairs of wavelet basis,” *IEEE Trans. Image Processing*, vol. 51, no. 12, pp. 3293-3294, Dec. 2003.
- (35) R. Yu, and H. Ozkaramanli, “Hilbert transform pairs of orthogonal wavelet bases : necessary and sufficient conditions,” *IEEE Transactions Signal Processing*, vol. 53, no. 12, pp. 4723-4725, Dec. 2005.
- (36) R. Yu, and H. Ozkaramanli, “Hilbert transform pairs of biorthogonal wavelet bases,” *IEEE Transactions Signal Processing*, vol. 54, no. 6, pp. 2119-2125, June, 2006.
- (37) N. G. Kingsbury, “The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement,” *Proceedings European Signal Processing Conference*, pp. 319-322, Rhodes, Sept. 1998.
- (38) A. Abbas, and T. Tran, “Fast approximations of the orthogonal dual-tree wavelet bases,” *Proceedings IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 605-608, Philadelphia, Mar. 2005.
- (39) W. T. Freeman, and E. H. Adelson, “The design and use of steerable filters,” *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, Sept. 1991.
- (40) E. P. Simoncelli, and W. T. Freeman, “The steerable pyramid : A flexible architecture for multi-scale transforms,” *IEEE Transactions Information Theory*, vol. 38, no. 2, pp. 587-607, Mar. 1992.
- (41) R. H. Bamberger, and M. J. T. Smith, “A filter bank for the directional decomposition of images : Theory and design,” *IEEE Transactions Signal Processing*, vol. 40, no. 4, pp. 882-893, Apr. 1992.
- (42) M. N. Do, and M. Vetteli, “Contourlets : A directional multiresolution image representation,” *Proceedings IEEE International Conference Image Processing*, vol. 1, pp. 357-360, Rochester, 2002.
- (43) J. -L. Starck, E. J. Candès, and D. L. Donoho, “The curvelet transform for image denoising,” *IEEE Transactions Image Processing*, vol. 11, no. 6, pp. 670-684, June 2000.

- (44) J. Liu, and P. Moulin, "Information-theoretic analysis of interscale and intracale dependencies between image wavelet coefficients," *IEEE Transactions Image Processing*, vol. 10, no. 11, pp. 1647–1658, 2001.
- (45) J. Yang, W. Xu, Q. Dai, and Y. Wang, "Image compression using 2D dual-tree discrete wavelet transform (DDWT)," *Proceedings IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 297–300, May 2007.
- (46) 高橋 良知, 渡辺 裕, "双対木複素ウェーブレットにおける疎表現の検討", 情報処理学会 AVM 研究会, 2007AVM58-06, Sep. 2007.
- (47) T H Reeves, and N G Kingsbury, "Prediction of coefficients from coarse to fine scales in the complex wavelet transform," *Proceedings IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, 6-9 June, 2000.
- (48) R. Anderson, N. Kingsbury, and J. Fauqueur, "Coarse level object recognition using interlevel products of complex wavelets," *Proceedings IEEE International Conference on Image Processing*, Genoa, Sept 2005.
- (49) M A Miller, and N G Kingsbury, "Statistical image modelling using interscale phase relationships of complex wavelet coefficients," *Proceedings IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, May 14-19, 2006, paper IMDSP-P13.8.
- (50) M A Miller, and N G Kingsbury, "Image estimation using interscale phase properties of complex wavelet coefficients," Accepted for publication in *IEEE Transactions on Image Processing*, July 2007.
- (51) T. H. Reeves, and N. G. Kingsbury, "R-D quantisation of complex coefficients in zerotree coding," *Proceedings 11th IEEE Workshop on Statistical Signal Processing 2001 (SSP2001)*, Singapore, August 2001.
- (52) J. Li, and S. Lei, "An embedded still image coder with rate-distortion optimization," *IEEE Transactions Image Processing*, vol. 8, no. 7, pp. 913–924, July, 1999.
- (53) T. Adeyemi, and M. Davies, "Sparse representations of images using overcomplete complex wavelets," *Proceedings 13th IEEE Workshop on Statistical Signal Processing 2001 (SSP2001)*, July 2005.
- (54) S. G. Mallat, and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

- (55) Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proceedings 27th Asilomar Conference Signals, Systems and Computers*, pp. 40–44, Nov. 1993.
- (56) P. S. Huggins, and S. W. Zucker, "Greedy basis pursuit," *IEEE Transactions Signal Processing*, vol. 2, pp. 3760–3772, July. 2007.
- (57) D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse over-complete representations in the presence of noise," *IEEE Transactions Information Theory*, vol. 52, pp. 6–18, Jan. 2006.
- (58) S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *Society for Industrial and Applied Mathematics (SIAM) Journal on Scientific Computing*, vol. 43, no. 1, pp. 129–159, 2001.
- (59) H. Bölcskei, and F. Hlawatsch, "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis", *Proceedings IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 2453–2456, April 1997.
- (60) N. G. Kingsbury, and T. H. Reeves, "Iterative image coding with overcomplete complex wavelet transforms," *Proceedings Conference on Visual Communications and Image Processing (VCIP)*, paper 1160, Lugano, July 8–11, 2003.
- (61) L. Mancera, and J. Portilla, "L0-norm-based sparse representation through alternate projections," *Proceedings IEEE International Conference Image Processing*, Oct. 2006.
- (62) D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Transactions Circuits and Systems*, CAS-25, 9, 1978.
- (63) D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse over-complete representations in the presence of noise," *IEEE Transactions Information Theory*, vol. 52, pp. 6–18, Jan. 2006.
- (64) 高橋 良知, 渡辺 裕, "複素ウェーブレットの疎表現における係数選択手法の検討", 情報処理学会全国大会, Mar. 2008.
- (65) P. J. Burt, and E. H. Adelson, "The laplacian pyramid as compact image code," *IEICE Transactions on Communications*, vol. COM-31, no. 4, pp. 532–540, April 1983.

- (66) 高橋 良知, 渡辺 裕, “複素ウェーブレット変換の疎表現における係数予測手法の検討”, 電子情報通信学会総合大会, Mar. 2008.
- (67) “講義資料：線形代数の基本定理 (東京大学 大学院情報理工学系研究科 石川並木小室研究室)”, <http://www.k2.t.u-tokyo.ac.jp/~koichi/Lecture/Pattern/2002/l4.pdf>
- (68) I. Daubechies, “Time–frequency localization operators : a geometric phase space approach,” *IEEE Transactions on Information Theory*, pp. 605–612, 1988.

付 録 A

値域と零空間

以下に、第 4.3.1 で述べた値域と零空間の詳細を述べる。

入力信号 $x \in \mathbb{R}^3$ と 3×4 の再構成行列 $\mathbf{R}$ があるとする。行列 $\mathbf{R}$ の列ベクトルは基底ベクトルに対応している。よって、 $\mathbf{R}$ は過完備基底から構成されている。このとき行列 $\mathbf{R}$ に対して、式 (A.1) で定義される値域と式 (A.2) で定義される零空間が存在する。

$$\text{Im}\mathbf{R} = \{x \in \mathbb{R}^3 | x = \mathbf{R}y, y \in \mathbb{R}^4\} \quad (\text{A.1})$$

$$\text{Ker}\mathbf{R} = \{y \in \mathbb{R}^4 | \mathbf{R}y = 0\} \quad (\text{A.2})$$

値域 $\text{Im}\mathbf{R}$ は $\mathbb{R}^3$ の部分空間、零空間 $\text{ker}\mathbf{R}$ は $\mathbb{R}^4$ の部分空間である。線形代数の基本定理より、 $\mathbf{R}$ の転置行列を $\mathbf{R}^T$ とすると、式 (A.3) が成立する⁽⁶⁷⁾。

$$\text{Ker}\mathbf{R} = (\text{Im}\mathbf{R}^T)^\perp \quad (\text{A.3})$$

これは $\text{Ker}\mathbf{R}$ の要素は、 $\mathbf{R}^T$ のすべての列ベクトルに直交することをあらわしている。これより、 $x = \mathbf{R}y$ の解 y は、転置行列 $\mathbf{R}^T$ の値域上への射影と $\mathbf{R}$ の零空間上への射影とに直交分解できる。

$$y = y_r + \omega \quad y_r^T \omega = 0, \quad y_r \in \text{Im}\mathbf{R}^T, \omega \in \text{Ker}\mathbf{R} \quad (\text{A.4})$$

ここで、例として次のような線型方程式 $x = \mathbf{R}y$ の解 y を考える。

$$x = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad (\text{A.5})$$

値域と零空間の定義より、転置行列 $\mathbf{R}^T$ の値域と $\mathbf{R}$ の零空間は式 (A.6) となる。

$$\text{Im}\mathbf{R}^T = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}, \quad \text{Ker}\mathbf{R} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix} \right\} \quad (\text{A.6})$$

よって、 $\mathbf{y}$ は 3 次元の値域 $\text{Im}\mathbf{R}^T$ と 1 次元の零空間 $\text{Ker}\mathbf{R}$ に属する係数から構成される。ここで、値域の各基底ベクトルに対応する係数を α, β, γ 、零空間の基底ベクトルに対応する係数を t とすると、 $\mathbf{y}$ は式 (A.7) となる。

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \alpha \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} + \beta \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \gamma \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + t \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \alpha + t \\ \beta + t \\ \gamma + t \\ \alpha + \beta + \gamma - t \end{bmatrix} \quad (\text{A.7})$$

今回の例では、入力信号 $\mathbf{x}$ を表現する値域の係数は $\alpha = -1, \beta = 1, \gamma = 3$ となる。そして、値域の定義より、この係数値を変更すると入力信号の再構成は不完全となり誤差が生じる。一方で、零空間の定義より、零空間の係数 t は入力信号に影響を与えないため任意の値をとることができる。

このように、過完備基底を用いた変換の解は値域と零空間から構成され、その数は無数に存在する。この無数の解のうち特定の性質をもつ解を求める手法が存在する。たとえば、フレーム手法 (MOF: Method Of Frame) と呼ばれる手法⁽⁶⁸⁾では、 l^2 ノルムを最小 ($\|\mathbf{y}\|_2 \rightarrow \min$) とする解を求める。これはピタゴラスの定理より、式 (A.4) において $\omega = 0$ のとき実現される。今回の例では、 $\alpha = -1, \beta = 1, \gamma = 3, t = 0$ の解 $\mathbf{y} = [-1 \ 1 \ 3 \ 3]^T$ がこれに相当する。しかし、MOF により求まる解は疎とならないことが知られている。疎表現問題では疎な解を求めるため l^2 ノルムではなく l^0 ノルムを最小 ($\|\mathbf{y}\|_0 \rightarrow \min$) とする解を求める。今回の例では、たとえば $t = -1$ とすることで解は $\mathbf{y} = [-2 \ 0 \ 2 \ 2]^T$ となり非ゼロ係数を 1 つ減少した疎な解 $\mathbf{y}$ が求まる。

図一覧

2-1 変換符号化の符号化器と復号化器の構成	7
2-2 2次元離散コサイン変換の基底ベクトル (文献 ⁽⁹⁾ 引用)	9
2-3 ミッドトレッド型量子化器 (文献 ⁽⁹⁾ 参照)	10
2-4 JPEG 2000 符号化アルゴリズム (文献 ⁽¹²⁾ 参照)	13
2-5 時間-周波数表現における領域の重複と冗長性	16
2-6 多重解像度解析における空間の入れ子構造	17
2-7 1次元離散ウェーブレット変換における分割フィルタバンク	19
2-8 1次元離散ウェーブレット変換における合成フィルタバンク	19
2-9 2分割フィルタバンク (文献 ⁽¹⁸⁾ 参照)	20
2-10 2次元離散ウェーブレット変換におけるフィルタバンク	22
2-11 2次元離散ウェーブレット変換の階層構造	23
2-12 自然画像に対する2次元離散ウェーブレット変換	24
2-13 離散ウェーブレット変換におけるシフト不変性の欠如	25
2-14 2次元離散ウェーブレット変換の方向分離特性	27
2-15 SPIHT 符号化器と復号化器 (文献 ⁽²³⁾ 参照)	28
2-16 サブバンド上における親子関係	29
2-17 ビットプレーン展開と量子化	31
2-18 離散ウェーブレット係数に対する SPIHT 符号化の符号化性能	32
3-1 ヒルベルト変換	34
3-2 解析信号の周波数表現	35
3-3 複素ウェーブレット変換の概念	36
3-4 1次元並列木複素ウェーブレット変換の分割フィルタバンク (文献 ⁽³³⁾ 参照)	38
3-5 1次元並列木複素ウェーブレット変換の合成フィルタバンク (文献 ⁽³³⁾ 参照)	39
3-6 2次元並列木複素ウェーブレット変換の分割フィルタバンク	44
3-7 自然画像に対する2次元並列木複素ウェーブレット変換	45
3-8 並列木複素ウェーブレット変換におけるシフト不変性	46
3-9 2次元並列木複素ウェーブレット変換の方向分離特性	48
3-10 並列木複素ウェーブレット変換の帯域分割特性	51
3-11 並列木複素ウェーブレット変換の位相変化	51
3-12 複素数空間の量子化セル	52
3-13 並列木複素ウェーブレット係数に対する SPIHT 符号化の性能	54

3-14 基底による信号表現	55
3-15 基底追跡の幾何的解釈	59
4-1 並列木複素ウェーブレット変換の疎表現係数列における符号化器と復号化器	61
4-2 並列木複素ウェーブレット変換における Noise Shaping 法 (文献 ⁽⁴⁶⁾ 参照)	62
4-3 Noise Shaping 法の収束過程	63
4-4 交互射影による l^0 ノルム最小化 (文献 ⁽⁶¹⁾ 参照)	64
4-5 提案手法と従来手法の係数選択の違い	66
4-6 パラメータ <i>Iteration</i> と非ゼロ係数へのエネルギー集約度 (文献 ⁽⁴⁶⁾ 参照)	69
4-7 パラメータ <i>NonZero</i> と非ゼロ係数へのエネルギー集約度 (文献 ⁽⁴⁶⁾ 参照)	69
4-8 提案手法と従来手法の非ゼロ係数へのエネルギー集約度 (文献 ⁽⁴⁶⁾ 参照)	69
4-9 実係数と複素係数の量子化	70
4-10 Lena における並列木複素ウェーブレット疎表現係数列の符号化特性	72
4-11 Barbara における並列木複素ウェーブレット疎表現係数列の符号化特性	73
4-12 Goldhill における並列木複素ウェーブレット疎表現係数列の符号化特性	74
4-13 ステップエッジに対する並列木複素ウェーブレット係数の絶対値応答	75
4-14 疎表現係数列における係数予測手法 (文献 ⁽⁶⁶⁾ 参照)	76
4-15 係数予測における重み係数と RMSE・PSNR 向上値 (文献 ⁽⁶⁴⁾ 参照)	77

表一覧

1-1 並列木複素ウェーブレット変換と離散ウェーブレット変換の違い	3
2-1 8×8 量子化テーブルの例 (文献 ⁽⁹⁾ 引用)	11
2-2 Daubechies 9/7 タップウェーブレットのフィルタ係数	23
3-1 並列木複素ウェーブレット変換のフィルタ係数	42
3-2 Lena における並列木複素ウェーブレット係数の相互情報量 (文献 ⁽⁴⁶⁾ 参照) .	50
3-3 Lena における離散ウェーブレット係数の相互情報量 (文献 ⁽⁴⁶⁾ 参照)	50
3-4 Barbara における並列木複素ウェーブレット係数の相互情報量	50
3-5 Barbara における離散ウェーブレット係数の相互情報量	50
4-1 Lena における疎表現係数列の絶対値の相互情報量 (文献 ⁽⁴⁶⁾ 参照)	68
4-2 Barbara における疎表現係数列の絶対値の相互情報量	68
4-3 Lena における係数予測手法の PSNR 達成率 (文献 ⁽⁶⁴⁾ 参照)	77
4-4 Barbara における係数予測手法の PSNR 達成率	77

研究業績

	題 名	発表年月	発表掲載誌	連名者
国内				
(1)	映像シーンに適応したスケラビリティ選択手法に関する検討	2005 年 11 月	PCSJ2005, P-5.13	高橋 良知 後藤 崇行 石川 孝明 渡辺 裕 富永 英義
(2)	映像シーンに適応したスケラビリティ選択手法	2006 年 3 月	信学春季全大, D-11-19	高橋 良知 石川 孝明 渡辺 裕 富永 英義
(3)	複素ウェーブレット変換を用いた画像符号化に関する一検討	2006 年 9 月	FIT2006, J-031	高橋 良知 河村 圭 渡辺 裕
(4)	複素ウェーブレットを用いた画像符号化における情報量削減手法に関する検討	2006 年 11 月	PCSJ2006, P-5.09	高橋 良知 河村 圭 渡辺 裕
(5)	双対木複素ウェーブレットにおける疎表現の検討	2007 年 9 月	情報処理学会 AVM 研究会 58-06	高橋 良知 渡辺 裕
(6)	複素ウェーブレットの疎表現における係数選択手法の検討	2008 年 3 月	情処全大	高橋 良知 渡辺 裕
(7)	複素ウェーブレット変換の疎表現における係数予測手法の検討	2008 年 3 月	信学春季全大	高橋 良知 渡辺 裕

P-5.13

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 との関係を、動画の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、 $L_3, H_3, H_2, H_2, \dots$ 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

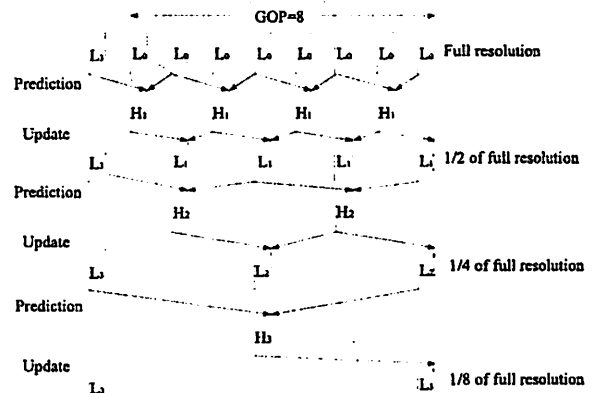


Figure 1: Structure of MCTF

3 予備実験

3.1 実験条件

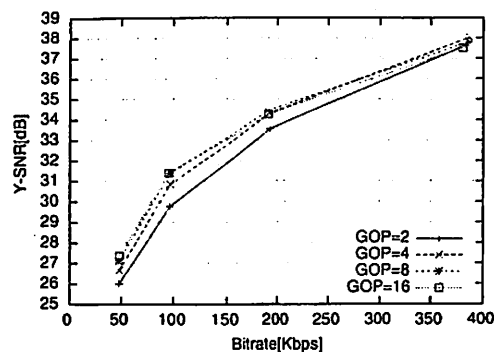
時間解像度の分割数と最高フレームレートでの SNR の関係を知るための予備実験を行った。予備実験は、SVC の参照ソフトウェア JSVM2.0 (Joint Scalable Video Model) [2] 上で行い、表 2 に示す実験条件に基づいた。なお、今回は MCTF の階層構造による影響を確認するために、Update の処理は省略した。

3.2 実験結果

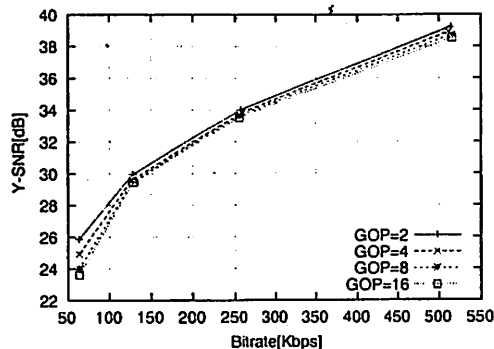
図 2 は、標準動画 Mobile&Calendar, Football について、GOP サイズと最高フレームレートにおける SNR の比較を行ったものである。Mobile では GOP サイズが 8 の時に SNR が最大となっている。一方で、Football では、GOP サイズが 2 の時に SNR が最大となっている。このことから、特定の GOP サイズで最高フレームレートでの映像の SNR を最大にすることができると分かる。さらに、その GOP サイズは動画の特徴に依存して異なる。他の標準動画については、Foreman, City は Mobile&Calendar と同様の特性を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

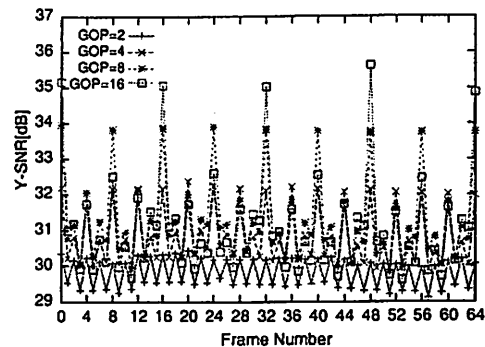


(a) Mobile

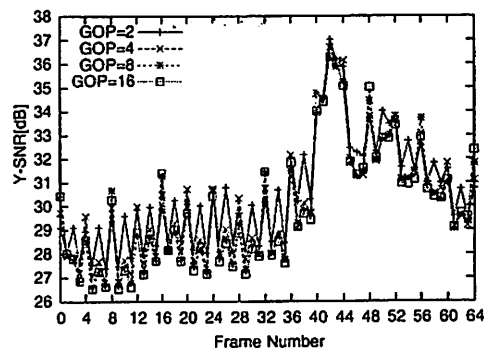


(b) Football

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 Information), SI(Spatial Information)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, Apr.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

早稲田大学理工学部 電子・情報通信学科 富永研究室
〒169-8555 東京都新宿区大久保 3-4-1 55-N-06-02
Phone: 03-5286-3365
E-mail: yoshitomo@tom.comm.uacsda.ac.jp

映像シーンに適応したスケーラビリティ選択手法

D-11-19

Adaptive Selection of Scalability Method for Video Scene

高橋 良知^{*1} 石川 孝明^{*2} 渡辺 裕^{*2} 富永 英義^{*1,*2}
Yoshitomo TAKAHASHI^{*1} Takaaki ISHIKAWA^{*2} Hiroshi WATANABE^{*2} Hideyoshi TOMINAGA^{*1,*2}^{*1} 早稲田大学 理工学部 電子・情報通信学科 ^{*2} 早稲田大学 国際情報通信研究センター
^{*1} Dept. of Elec. Info. and Comm. Eng., WASEDA Univ. ^{*2} Global Info. and Tele. Institute, WASEDA Univ.

1 はじめに

近年、ネットワーク帯域、および受信端末の解像度の多様化により、ワンソース・マルチユースの映像配信を実現するための技術として、映像のスケーラビリティ機能が注目されている。現在 MPEG では H.264/MPEG-4 AVC をベースとした SVC (Scalable Video Coding) の標準化が進められている [1]。SVC では、空間・時間・SNR スケーラビリティを提供することができる。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 [2]。

そこで、本稿では、時間・SNR スケーラビリティに着目し、SVC におけるスケーラビリティの粒度と符号化効率の関係に基づき、SNR スケーラビリティを提供する FGS (Fine Granular Scalability) 機能を切り替えることによる主観品質の向上のための手法を検討する。

2 スケーラビリティと符号化効率

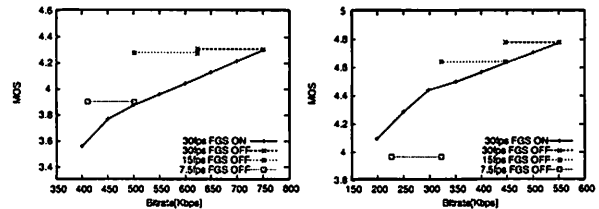
SVC は、時間スケーラビリティを実現するために MCTF (Motion Compensated Temporal Filtering) を用いる。MCTF は、時間方向のサブバンド分割を繰り返して実行することで、階層構造を構成し、時間スケーラビリティを実現している。MCTF では、GOP サイズに応じて、取得可能なフレームレートが変化する。一般に、 2^N ($N = 1, 2, 3, \dots$) の GOP サイズでは、 N 階層の時間解像度が得られる。動画の時間複雑度や動き補償予測の精度に応じて、フルデコードの際の平均 PSNR の最大値を与える GOP サイズが異なることが分かっている [3]。

また、SNR スケーラビリティの実現には FGS が用いられる。SVC は、Cyclic Block Coding と呼ばれる符号化を行うが、これは基本レイヤと拡張レイヤへ分割するために、MPEG-4 の Bitplane Coding と同様に、非 FGS 符号化と比較し、約 1~1.5[dB] 程の符号化効率の低下が見られる。

3 検討事項

FGS を用いたスケーラビリティ制御では、粒度の高い SNR スケーラビリティ制御が可能であるが、非 FGS に対し PSNR の低下が発生する。一方で、MCTF を用いたスケーラビリティ制御では、粒度の低い時間スケーラビリティ制御となるが、高い PSNR を維持することが可能である。そこで、動画の時間方向と SNR 方向の主観に対する重要性に基づき、これらのスケーラビリティを切り替えることで、主観品質の向上を検討する。

文献 [4] では、時間方向の主観品質を MOS_t 、空間方向の主観品質を MOS_s とし、それぞれ的主観品質の推定式が提案され、これらの推定式に基づき、時間・SNR スケーラビリティの組み合わせ手法が検討されている。 MOS_t の推定の特徴量としては動きベクトルの最大値が用いられ、 MOS_s の推定の際の特徴量としては Sobel フィルタによるエッジの統計量が用いられている。本稿においても、同様の特徴量を用いているが、文献 [4] では、MPEG-4 をベースとした動き補償推定を基本とし



(a) FOOTBALL

(b) HARBOUR

図1 スケーラビリティの実現方法と MOS の関係

ているのに対し、本稿は H.264 を基本とする SVC 上で検討を行っているため、可変ブロックサイズや動き補償の参照フレーム間距離の変化などが生じる。そこで、今回は、可変ブロックについては、相関の強い動きベクトルを有するブロックのみを利用し、参照フレーム間の距離については、距離に応じた値で正規化を行う。

4 結果

図1は、標準動画 FOOTBALL, HARBOUR について、スケーラビリティを時間・SNR のいずれか一方により実現した場合の、推定 MOS を示している。なお、GOP サイズは、フルデコードの際の PSNR が最も高くなるサイズとしている。FOOTBALL などの時間複雑度の大きいシーケンスでは、フレームレートの低下が主観品質の低下に与える影響が小さく、かつ、低フレームレートにすることにより1フレームあたりに費やすビット量が増加する。このときの PSNR の向上が大きいため、時間スケーラビリティによる制御では、7.5[fps]においても FGS を用いた 30[fps] に比べて高い MOS が得られたと考えられる。

5 まとめと今後

スケーラビリティの粒度と符号化効率の関係に着目し、動画の特性に応じてスケーラビリティを時間・SNR のいずれか一方により実現することによる主観品質の向上を検討した。今後は、実際に主観評価実験を行い検討手法の有効性を確認するとともに、SNR・時間スケーラビリティの切り替え方法について検討を行う。

参考文献

- [1] J. Reichel et al., "Scalable Video Coding - Working Draft 4," Joint Video Team(JVT) of ISO/IEC MPEG & ITU-T VCEG 16th Meeting, Doc. JVT-Q201, Oct. 2005.
- [2] 木本, "MPEG におけるスケーラブルビデオ符号化の標準動向," 情報学, AVM-48, P-55.60, 2004.
- [3] 高橋, 石川, 後藤, 渡辺, "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," PCSJ2005.
- [4] 富田, 吉田, "要求された主観品質を実現する動画像符号化のレート制御法," PCSJ2005.

複素ウェーブレット変換を用いた画像符号化に関する一検討 A Study on Image Coding with Complex Wavelet Transform

高橋 良知†

河村 圭†

渡辺 裕†

Yoshitomo TAKAHASHI

Kei KAWAMURA

Hiroshi WATANABE

1. まえがき

離散ウェーブレットを用いた2次元多重解像度解析(2D-DWT: Discrete Wavelet Transform)は、冗長性を持たない変換であるため、画像符号化に広く利用されている。一方で、シフト不変性の欠如は、HH成分における $\pm 45^\circ$ 方向のエッジの混在、という画像品質の劣化に影響を与える特性を有する。シフト不変性の欠如は、画像中のエッジ付近のウェーブレット係数がLevel間で大きく変化する原因となる。また、HH成分における方向性の混在は、低ビットレートにおけるチェッカーボードアーティファクトの原因となる[1]。

シフト不変性問題を解決する手法として、複素ウェーブレットが提案されている[2]。複素ウェーブレットを用いた多重解像度解析は、 m 次元の信号に対し $2^m:1$ という冗長性を持たせることで、近似的にシフト不変性が成立し、 $\pm 15^\circ$ 、 $\pm 45^\circ$ 、 $\pm 75^\circ$ の6方向の画像エッジを独立に持つ高域サブバンドを生成する。

著者らは、上記の特性を利用し、複素ウェーブレットを画像符号化へ適用する検討を行なっている。本稿では、複素ウェーブレットのシフト不変性を利用し、補間による高域サブバンドのエントロピー削減手法について検討を行う。

2. 2次元複素ウェーブレット多重解像度解析

2.1 複素ウェーブレット

画像圧縮に用いる離散ウェーブレットの大部分は実数型ウェーブレットであるが、この実数型ウェーブレットの単一位相がシフト不変性の欠如の原因である。そこで、このシフト不変性の欠如を解決する手法として、式(1)のように、2種類の異なる実数型ウェーブレット $\psi_r(t)$ 、 $\psi_i(t)$ を実数部、虚数部に配置した複素ウェーブレットが提案されている[2]。また、 $\psi_r(t)$ 、 $\psi_i(t)$ に対応するスケーリング関数を $\phi_r(t)$ 、 $\phi_i(t)$ とすると、式(2)により、複素ウェーブレットのスケーリング関数が定義される。

$$\psi_c(t) = \psi_r(t) + j\psi_i(t) \quad (1)$$

$$\phi_c(t) = \phi_r(t) + j\phi_i(t) \quad (2)$$

実数部、虚数部に配置する実数型ウェーブレットには様々な組み合わせが考えられ、ヒルベルト変換ペア[3]や双直交Splineウェーブレットを用いたRI-Splineウェーブレット[4]などが提案されている。

2.2 複素ウェーブレット多重解像度解析の構造

複素ウェーブレットによる2次元多重解像度解析(2D-CWT: Complex Wavelet Transform)の処理フローを図1に示す。入力画像に対し、実数部、虚数部、それぞれのウェーブレットを用いて、行方向、列方向の順に1次元多重解像度解析を行なうことで実現される。2D-CWTでは、1つのLevelについて、12種類の高域サブバンドと4種類の低域サブバンドの計16種類のサブバンドが生成されるため、4:1の冗長性を持つ。

2D-CWTの後に得られた12種類の高域サブバンドについて、式(3)、(4)による変形を行い、得られたサブバンドを式(5)、(6)のように、実数部、虚数部に配置することで、2つの複素数を得られる。ここで、 $i=1,2,3$ であり、 $\phi_{1,i}$ 、 $\phi_{2,i}$ は、図1における $R \times R$ 、 $I \times I$ の3つの高域サブバンド、 $\phi_{3,i}$ 、 $\phi_{4,i}$ は、 $R \times I$ 、 $I \times R$ の3つの高域サブバンドである。

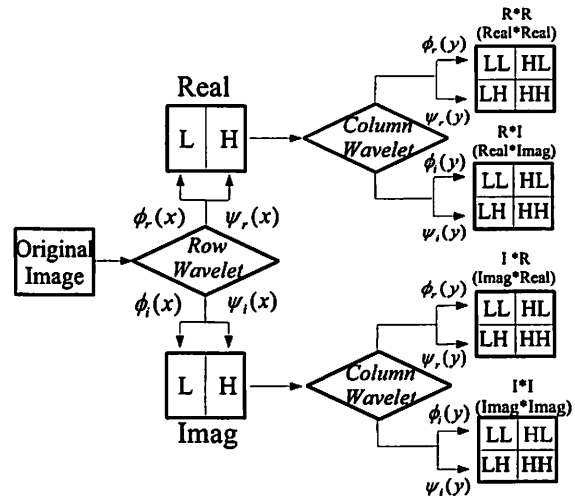


図1: 2D-CWTにおけるLevel1への分解

$$\begin{cases} \psi_i^{+,r}(x,y) = \frac{1}{\sqrt{2}}(\phi_{1,i}(x,y) - \phi_{2,i}(x,y)) \\ \psi_i^{-,r}(x,y) = \frac{1}{\sqrt{2}}(\phi_{1,i}(x,y) + \phi_{2,i}(x,y)) \end{cases} \quad (3)$$

$$\begin{cases} \psi_i^{+,i}(x,y) = \frac{1}{\sqrt{2}}(\phi_{3,i}(x,y) - \phi_{4,i}(x,y)) \\ \psi_i^{-,i}(x,y) = \frac{1}{\sqrt{2}}(\phi_{3,i}(x,y) + \phi_{4,i}(x,y)) \end{cases} \quad (4)$$

$$\psi_i^+(x,y) = \psi_i^{+,r}(x,y) + j\psi_i^{+,i}(x,y) \quad (5)$$

$$\psi_i^-(x,y) = \psi_i^{-,r}(x,y) + j\psi_i^{-,i}(x,y) \quad (6)$$

2つの複素数のうち、 $\psi_i^+(x,y)$ の絶対値は、 15° 、 45° 、 75° 方向の画像エッジ付近で大きい値を持ち、 $\psi_i^-(x,y)$ は、 -15° 、 -45° 、 -75° 方向の画像エッジ付近で大きい値を持つ。

また、式(5)、(6)により得られた2つの複素数において、実数部、虚数部、それぞれ独立ではシフト不変性は成立しないが、複素数の絶対値に関して近似的にシフト不変性は成立する。

3. 複素ウェーブレットによる画像符号化

3.1 符号量の削減方法

2D-CWTは4:1の冗長性を持つ変換であるため、このまま符号化へ適用すると符号量が増加する。そこで、このような冗長信号に基づく信号圧縮は、解析信号の特性を良く表す係数を順に抽出してゆくことで行なわれる[5]。

この手法を複素ウェーブレットに適用したNoise Shaping法が提案されている[6]。Noise Shapingは、複素ウェーブレット変換後の係数に対し、係数間のエネルギーの移動を行なうことで、非ゼロ係数を減少させる手法である。

次節で提案する高域サブバンドのレベル間相関の除去は、このような手法と組み合わせて用いることで、さらにエントロピーを削減することを想定している。

†早稲田大学大学院 国際情報通信研究科
Graduate School of Global Information and Telecommunication
Studies, Waseda University.

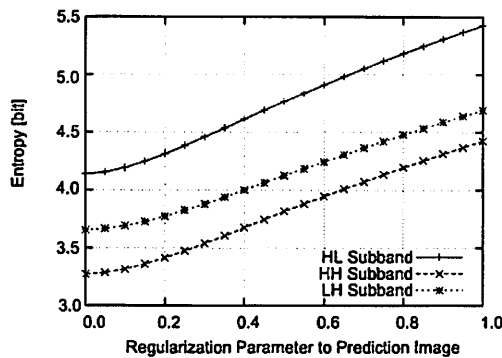


図 2: 2D-DWT の Level1 における高域サブバンドのエントロピー

表 1: 2D-CWT 後の Level1 高域サブバンドのエントロピー

Subband	15°	45°	75°	-15°	-45°	-75°	Total
Predict	3.60	2.46	2.76	3.70	2.72	3.14	18.38
Not Predict	4.50	3.00	3.43	4.61	3.23	3.80	22.57

表 2: 2D-DWT 後の Level1 高域サブバンドのエントロピー

Subband	0°	± 45°	90°	Total
DWT	4.14	3.28	3.66	11.08

3.2 離散ウェーブレットにおけるレベル間相関

実数型ウェーブレットにより得られた高域サブバンドは、シフト不変性が欠如しているため高域サブバンドの Level 間の相関性が低いと考えられる。図 2 は、2D-DWT の後、Level2 の高域サブバンドを線形補間して得られた予測バンドと Level1 の高域サブバンドの間の差分係数のエントロピーを示している。入力画像は Lena (512×512)、フィルタは Daubechies 8,8 フィルタを用いた。横軸は、予測係数のエネルギーを正規化する係数である。それぞれの高域サブバンドにおいて、サブバンド間の予測が適切になされないため差分係数のエントロピーは単調に減少している。そして、予測を行わない場合にエントロピーが最小となる。

3.3 複素ウェーブレットにおけるレベル間相関

式 (5), (6) の絶対値において、Level j のウェーブレット係数の分布は、画像中のエッジ中心からの距離 k に応じて、式 (7) のガウス分布で近似できることが知られている [7]。

$$L_j(k) = K e^{-\frac{1}{2}(\frac{k}{\sigma_n})^2} \quad (7)$$

ここで、 K はガウス関数の頂点の高さ、 σ_n はガウス関数の分散である。この近似は Level に関係なく成立するが、 K , σ_n はともに高い Level ほど大きくなる。2D-CWT 後の高域サブバンドは、この特性があるため Level j の高域サブバンドを Level $j+1$ の高域サブバンドから予測することが可能である。ここでは、Level $j+1$ の高域サブバンドを 1:2 にアップサンプリングし、これにローパスフィルタを施したものを予測バンドとする。これと Level j の高域サブバンドとの差分を計算する。ローパスフィルタは、

- 平均値フィルタ (線形補間)
- ガウシアンフィルタ (分散: 1.7, 窓長: 11)
- 複素ウェーブレットの実数部の再構成スケーリング関数の 3 種類を比較する。予測は複素数の絶対値で行うが、再構成時には、複素数の実数部と虚数部の各々の値が必要となる。そこで、複素数の絶対値で求めた残差係数を実数部と虚数部に分配する。このとき、予測前の位相を保存するように、実数部、虚数部へ残差係数を分配することで、完全再構成を保持することができる。また、複素数の絶対値で得られた残差係数は、正負のいずれかの符号が発生するため、残差係数には符号フラグを付加する。

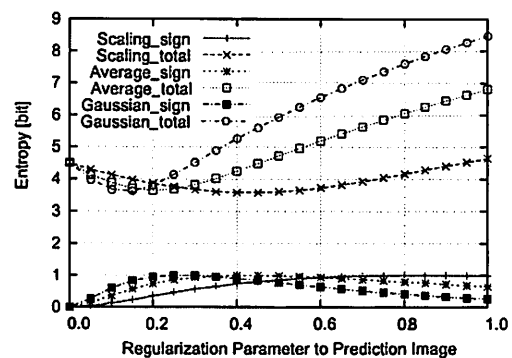


図 3: 2D-CWT の Level1 における 15° 方向高域サブバンドのエントロピー

4. 実験結果・考察

本稿では、複素ウェーブレットとして、RI-Spline ウェーブレット [4] を用いて実験を行なった。15° 方向の高域サブバンドの実験結果を図 3 に示す。図 3 は、3 種類のローパスフィルタについて、高域サブバンド間の予測後の残差係数の符号フラグのエントロピーと、実数部、虚数部の残差係数と符号フラグの合計のエントロピーを表している。

いずれのローパスフィルタを用いた場合でも、合計のエントロピーが極小値を持つ。これは、適切な正規化係数を用いて高域サブバンド間の予測を行なうことで、高域サブバンド間の相関を除去し、エントロピーを削減できることを示している。図 3 では、合計のエントロピーの最小値は、ローパスフィルタの種類によらずほぼ同じである。しかし、符号フラグを除いた残差係数のみのエントロピーは、スケーリング関数を用いた場合が最も小さいため、Noise Shaping と組み合わせるとときに他の 2 つのローパスフィルタよりもエントロピーを小さくできると考えられる。また、最小のエントロピーを与える正規化係数の値は、用いるローパスフィルタの種類により異なるが、同一のローパスフィルタの場合、全方向の高域サブバンドにおいてほぼ同じ値となる。

2D-CWT 後の Level1 の全高域サブバンドにおける予測の有無によるエントロピーの違いを表 1 に示す。高域サブバンド間の予測を行なうことで、全体で約 4.0[bit] のエントロピーを削減できる。また、2D-DWT 後の Level1 の全高域サブバンドのエントロピーを表 2 に示す。

5. むすび

本稿では、複素ウェーブレットのシフト不変性を利用し、補間を用いた高域サブバンド間予測を行なうことで、高域サブバンドのエントロピーを削減する手法について検討を行なった。実験より、予測サブバンドを用いることで、高域サブバンド間の相関を除去し、エントロピーを削減できることを確認した。

今後は、本手法を Noise Shaping と組み合わせることで、離散ウェーブレット変換を用いた画像符号化方式よりも優れた符号化効率を実現することを目指す。

参考文献

- [1] I.W. Selesnick, R.G. Baraniuk, and G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," IEEE Signal Processing Magazine, pp. 123-151, Volume 22, No. 6, Nov. 2005.
- [2] N.G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," Applied Computational Harmonic Anal., Vol. 10, no. 3, pp. 234-253, May 2001.
- [3] I.W. Selesnick, "Hilbert Transform Pairs of Wavelet Bases," IEEE Signal Processing Letters, 8-6, pp. 170-173, 2001.
- [4] H. Kawabata, H. Toda, Z. Zhang, and H. Fujiwara, "A new complex wavelet transform by using RI-spline wavelet," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Montreal, vol. 2, pp. 937-940, May 2004.
- [5] S.G. Mallat, Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Processing, Vol. 41, pp. 3397-3415, Dec. 1993.
- [6] T.H. Reeves and N.G. Kingsbury, "Overcomplete image coding using iterative projection-based noise shaping," ICIP 02, Rochester, NY, Sept 2002.
- [7] T.H. Reeves and N.G. Kingsbury, "Prediction of Coefficients from Coarse to Fine Scales in the Complex Wavelet Transform," in IEEE Proc. Int. Conf. Acoust., Speech, Signal Processing, 2000.

P-5.09

複素ウェーブレットを用いた画像符号化における情報量削減手法に関する検討 A Study on Reduction of Information in Image Coding Using Complex Wavelet

高橋 良知

河村 圭

渡辺 裕

Yoshitomo TAKAHASHI Kei KAWAMURA Hiroshi WATANABE

早稲田大学大学院 国際情報通信研究科

Graduate School of Global Information and Telecommunication Studies, Waseda University.

Abstract: CWT (Complex Wavelet Transform) provides approximate shift invariance and good directional selectivity, unlike DWT (Discrete Wavelet Transform). However, it is an overcomplete transform with $2^m : 1$ for m -dimensional signals. To achieve image coding using CWT, this redundancy should be reduced by the characteristic of CWT. "Noise Shaping" is an efficient redundancy reduction method, which generally increases a noise reduction in oversampled Filter Banks. Complex wavelet coefficients are modified with nearly loss-less quality by the NS method. However, modification of coefficients depends on a threshold of the NS method. In this paper, we clarify a relationship between the modification and the threshold.

1 はじめに

離散ウェーブレットを用いた2次元多重解像度解析(以下, DWT: Discrete Wavelet Transform)は, 冗長性を持たない変換である。これに対し, 近年, 複素ウェーブレットと呼ばれるウェーブレットが提案されている[1]。複素ウェーブレットはシフト不変性の成立, 画像中のエッジを6方向に分離可能, という離散ウェーブレットとは異なる特性を有している。また, 複素ウェーブレットは m 次元の信号に対し, $2^m : 1$ の冗長性を持つ。そこで, 複素ウェーブレットを用いた画像符号化を実現する場合, 上記の特性を利用し, 冗長性を削減していく必要がある。

本稿では, 複素ウェーブレット変換(以下, CWT: Complex Wavelet Transform)後の非ゼロ係数を効率的に削減することを目的とする。そこで, Noise Shaping(以下, NS)法による係数変化と閾値の関係について検討を行う。

2 複素ウェーブレットを用いた画像符号化

2.1 複素ウェーブレット変換

画像圧縮などに用いる実数型ウェーブレットに対し, 複素ウェーブレットは, 式(1)のように, 2種類の異なる実数型ウェーブレット $\psi_r(t)$, $\psi_i(t)$ を実数部, 虚数部に配置したものとなる。

$$\psi_c(t) = \psi_r(t) + j\psi_i(t) \quad (1)$$

$$\phi_c(t) = \phi_r(t) + j\phi_i(t) \quad (2)$$

$\psi_r(t)$, $\psi_i(t)$ に対応するスケーリング関数を $\phi_r(t)$, $\phi_i(t)$ とすると, 式(2)により, 複素ウェーブレットのスケーリング関数が定義される。この複素ウェーブレットを用いた2D-CWTの処理フローを図1に示す。2D-CWTは, 1つのレベルについて, 12種類の高域サブバンドと4種類の低域サブバンドの計16種類のサブバンドが生成されるため, 4:1の冗長性を持つ変換となる。

2.2 Noise Shaping 法

冗長フィルタバンクを用いた変換は, 変換後のデータ数が増加するため雑音に対する感度が低いという特性がある[2]。この特性を利用し, 変換・逆変換の過程で雑音を低減させる手法がNS法である。これを複素ウェーブレット変換に適用するための構成図を図2に示す[3]。NS法は繰り返しを伴う処理であるが, 一定回数の繰り返しで必ず収束する。この操作により, 閾値以下の複素ウェーブレット係数は次第に小さくなり, 閾値以上の複素ウェーブレット

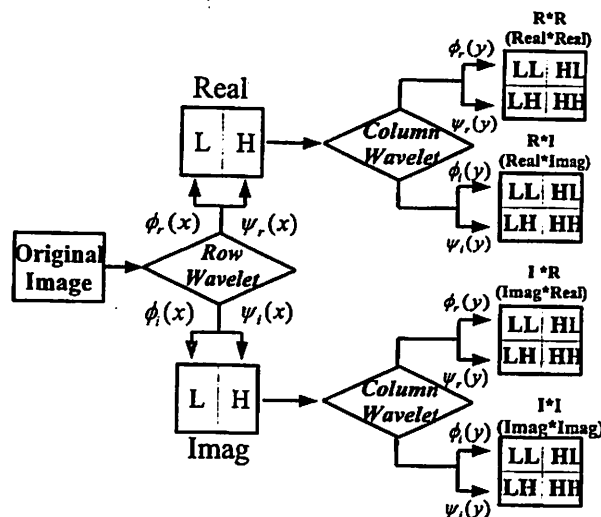


Figure 1: 2D-CWT in Level 1

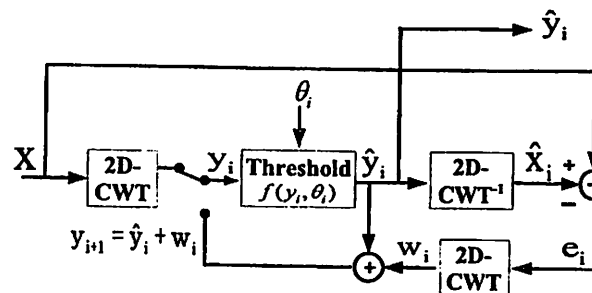


Figure 2: Noise Shaping to 2D-CWT

係数は次第に大きくなる, という効果が得られる。また, 再構成画像はほぼロスレスとなる。

3 予備実験

NS法を複素ウェーブレットに適用した際の非ゼロ係数とPSNRの関係を図3に示す。非ゼロ係数は最大値のウェーブレット係数から順番に指定個数まで抽出した。画像はLena (256 × 256) を使用し, 複素ウェーブレットとしてRI-Spline Wavelet[4], 離散ウェーブレットとしてDaubechies 9/7 Filterを用いた。また, 図2における閾値関数 $f(y_i, \theta_i)$ として, 以下のハード閾値を用いた。

$$\hat{y}^{2k-1} + j\hat{y}^{2k} = \begin{cases} 0 & \text{if } |y^{2k-1} + jy^{2k}| < \theta \\ y^{2k-1} + jy^{2k} & \text{otherwise} \end{cases} \quad (3)$$

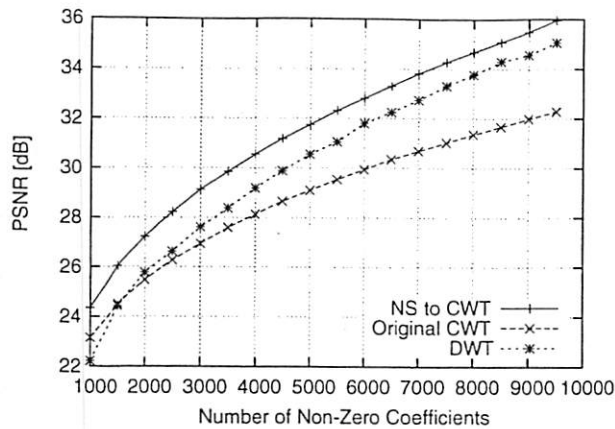
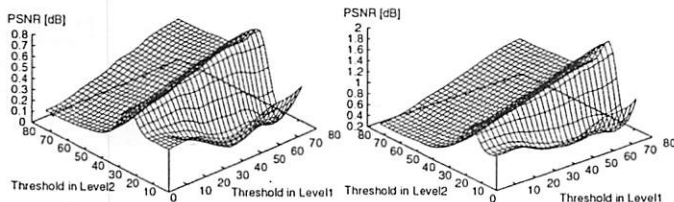


Figure 3: PSNR in Non-Zero Coefficients



(a) 400 Non-Zero Coef.

(b) 1200 Non-Zero Coef.

Figure 4: Relation between PSNR and Threshold in Level1

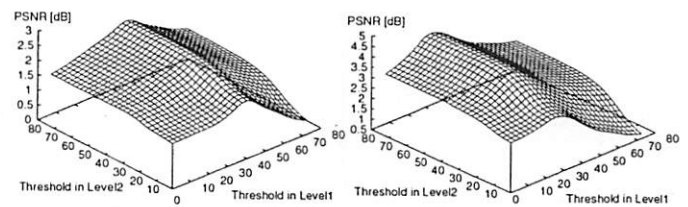
ここで、 y^{2k-1} 、 y^{2k} はそれぞれ k 番目の複素係数の実数部、虚数部である。図 3 より、NS 後の複素ウェーブレットでは、離散ウェーブレットと同数の非ゼロ係数において、約 1.0[dB]–2.0[dB] 程度の PSNR の向上がみられる。しかし、図 3 は実験的に求めた理想的な閾値関数 $f(y_i, \theta_i)$ を用いた場合であり、閾値関数 $f(y_i, \theta_i)$ により、同数の非ゼロ係数から再構成される画像の PSNR は変化する。そこで、同数の非ゼロ係数で最大の PSNR を与えるような閾値関数 $f(y_i, \theta_i)$ を与える必要がある。

4 検討手法

予備実験では各レベルにおいて同じ閾値を用いているが、レベルごとに画像エッジに対するウェーブレット係数の重要度が異なると考えられる。ここでは、レベルごとに可変の閾値を用いて、レベル 2 までの 2D-CWT について検討を行った。図 4、5 は、それぞれ、特定レベルから 400[個]、1200[個] の非ゼロ係数を抽出した場合の各レベルの閾値と PSNR の向上度合いを表したものである。

レベル 1 から抽出する非ゼロ係数は、レベル 2 に対する閾値に依存した極大値を持ち、レベル 2 から抽出する非ゼロ係数は、レベル 1 に対する閾値に依存した極大値を持つことがわかる。また、極大値を与える閾値は取り出す非ゼロ係数の個数に依存し規則的に変化する、という特性がある。このような特性は画像によらず成立するが、極大値を与える閾値は画像の特徴により変化する。

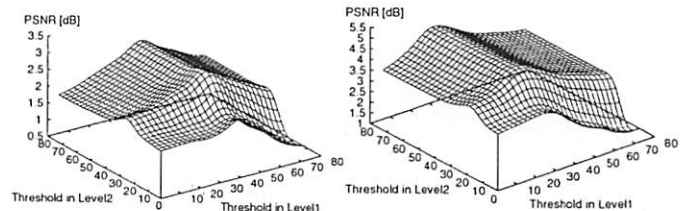
図 6 は、レベル 1、2 から同数個の非ゼロ係数を抽出した場合の PSNR の向上度合いを表したものである。各レベルにおける特定の閾値の組み合わせで極大値を形成するグラフが得られる。このとき、最大の PSNR を与えるときの閾値は各レベルで同値ではないことから、レベルごとに可変の閾値を用いる手法は有効であると考えられる。



(a) 400 Non-Zero Coef.

(b) 1200 Non-Zero Coef.

Figure 5: Relation between PSNR and Threshold in Level2



(a) 800 Non-Zero Coef.

(b) 2400 Non-Zero Coef.

Figure 6: Relation between PSNR and Threshold in Level1&2

実際は、レベルに無関係に非ゼロ係数を最大値から順番に抽出するため、各レベルで取り出される係数の個数は同じではないが、上記の特性は保持されるため、同数の非ゼロ係数で最大の PSNR を与える閾値が存在することになる。

5 おわりに

本稿では、複素ウェーブレットを用いた画像符号化について検討を行った。特に、非ゼロ係数を効率的に削減する Noise Shaping 法に着目し、レベルごとの閾値と係数変化の関係について検討を行った。その結果、収束後の非ゼロ係数と閾値の間には画像の特徴によらない一定の関係があることがわかった。

今後の課題として、画像の特徴より最適な閾値を自動で決定する方法や複素ウェーブレットに適したエントロピー符号化の検討などがあげられる。

参考文献

- [1] I.W. Selesnick, *et al.*: "The Dual-Tree Complex Wavelet Transform," IEEE Signal Processing Magazine, pp. 123–151, Nov. 2005.
- [2] H Bölcskei, *et al.*: "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis," IEEE ICASSP, pp. 2453–2456, April 1997.
- [3] N. G. Kingsbury, *et al.*: "Redundant Representation with Complex Wavelets: How to Achieve Sparsity," IEEE ICIP, Sept. 2003.
- [4] H. Kawabata, *et al.*: "A new complex wavelet transform by using RI-spline wavelet," IEEE ICASSP, pp. 937–940, May 2004.

早稲田大学大学院 国際情報通信研究科
〒367-0035 埼玉県本庄市西富田大久保山 1011
Phone: 0495-24-6143, Fax: 0495-24-6645
E-mail: yoshitomo@tom.comm.waseda.ac.jp

双対木複素ウェーブレットにおける疎表現の検討

高橋 良知[†] 渡辺 裕[†]

[†] 早稲田大学大学院 国際情報通信研究科
〒367-0035 埼玉県本庄市西富田 1011
E-mail: tyoshitomo.taka@toki.waseda.jp

あらまし 双対木複素ウェーブレット変換は、多次元信号において精度の良い方向選択性を持ち、シフト不変性が成り立つ。また、その2次元変換は変換係数の個数が4倍に増加する冗長変換である。冗長変換では少数の基底で対象物を表現する疎表現を求めることができる。本稿では、以下の2つを考察する。1つは、複素ウェーブレットにおいて疎表現を求める Noise Shaping と呼ばれる手法が L0 ノルムベースの疎表現近似手法と等価であることを示し、Noise Shaping の疎表現の近似性能を考察する。もう1つは、Noise Shaping の反復処理を制御するパラメータと求める疎表現の関係を Energy Compaction と相互情報量を用いて考察する。等価性より、従来とは異なる係数選択手法を提案し、提案手法において高い Energy Compaction を得られることを確認する。

キーワード 複素ウェーブレット, Noise Shaping, 疎表現, 交互射影, エネルギー集中性, 相互情報量

A Study on Sparse Representation of Dual-Tree Complex Wavelet

Yoshitomo TAKAHASHI[†] and Hiroshi WATANABE[†]

[†] Graduate School of Global Information and Telecommunication Studies, Waseda University
1011 Nishi-Tomida Honjyo-shi Saitama 367-0035 Japan
E-mail: tyoshitomo.taka@toki.waseda.jp

Abstract The Dual-Tree Complex Wavelet Transform (DT-CWT) has good directional selectivity in multiple dimensions and its subband responses are approximately shift invariance. For two dimensional signals, this transform is an overcomplete transform with 4:1 redundancy. In overcomplete transforms, we can find sparse representation which is the one that uses as few basis as possible to represent an object. In this paper, we describe the following two facts. First, it is described that Noise Shaping method on DT-CWT to find sparse representation is equivalent to the approximate method of sparse representation based on L0 norm. Second, A relation between parameters of Noise Shaping and sparse representations by energy compaction and mutual information is clarified. Then, an improved selection of coefficients in Noise Shaping is proposed. Good Energy Compaction on proposed method is validated.

Key words Complex Wavelet, Noise Shaping, Sparse Representation, Alternating Projection, Energy Compaction, Mutual Information.

1. はじめに

JPEG 2000 などの画像符号化で用いるウェーブレットは最大間引きフィルタバンクで構成されることが一般的である。このフィルタバンクを用いた離散ウェーブレット変換 (DWT : Discrete Wavelet Transform) は、画像処理に不適切な2つの特性が知られている。1つはシフト不変性の欠如である。これは入力信号の微小シフトが各レベルの DWT 係数のエネルギーを大きく変えてしまう現象である。もう1つは、2次元画像において、対角方向の画像エッジを同一のサブバンド内に含んでしまう現象である。このため $+45^\circ$ と -45° の2方向を中心

とした画像エッジを分離して扱うことができず、画像符号化において量子化操作などによりチェッカーボードアーティファクトが発生し、主観品質が低下する。

これらの欠点を改善するウェーブレットとして、双対木構造を用いた複素ウェーブレット変換 (DT-CWT : Dual-Tree Complex Wavelet Transform) が提案されている [1]。DT-CWT はシフト不変性が近似的に成立し、 $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ の6方向を中心とした画像エッジを異なるサブバンドへ分離することができる [2]。また DT-CWT は冗長な変換であり、2次元変換では入力画像の4倍の変換係数が発生する。DWT のような非冗長な変換は対象物をただ1通りの基底の線形結合で表現

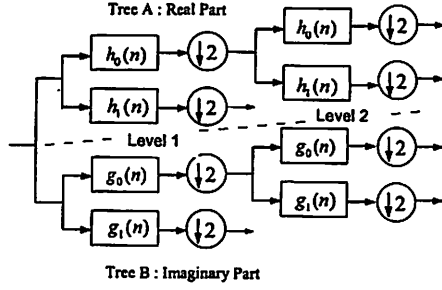


図 1 Analysis FB for 1D DT-CWT

するが、DT-CWT のような冗長な変換では対象物を表現する線形結合の組み合わせが無数に存在する。この中からできるだけ少数の基底で対象物を表現する疎表現を求め、これを画像処理に適用する研究がなされている。また近年、冗長基底を用いた画像符号化方式も提案され、特に低ビットレートにおける有効性が示されている [3]。

Kingsbury らは、複素ウェーブレットにおいて疎表現を求める Noise Shaping と呼ばれる手法を提案している [4]。しかし、この手法は一般的な疎表現近似手法 [5], [6] と異なるアプローチであり、これらの手法との関係性の考察が不十分である。また、Noise Shaping は反復処理であり、反復条件に基づき求まる疎表現の特性が変化するというような考察はされていない。

そこで、本稿では、Noise Shaping が Maceran の提案する L0 ノルムベースの疎表現近似手法 [7] と等価であることを示し、一般的な疎表現近似手法の観点から考察する。さらに、この等価性に基づき、Noise Shaping において従来とは異なる係数選択手法を提案し、提案手法において高い Energy Compaction が得られることを示す。また、Noise Shaping の反復処理を 2 つをパラメータで定義し、これらのパラメータと疎表現の特性を Energy Compaction と相互情報量を用いて関連付ける。

以下、2 章では複素ウェーブレットについて述べる。3 章では Noise Shaping について述べ、反復処理を制御するパラメータを定義する。4 章では一般的な疎表現近似手法と Mancera らが提案した L0 ノルムベースの疎表現近似手法について述べる。5 章では Noise Shaping と Mancera らの手法が等価であることを示し、その等価性に基づく新たな係数選択手法を提案する。6 章では提案手法において高い Energy Compaction が得られることを確認し、3 章で定義したパラメータと求まる疎表現との関係を Energy Compaction と相互情報量を用いて評価する。7 章は本稿のまとめである。

2. 双対木複素ウェーブレット

2.1 複素ウェーブレット

DWT の問題点は実数型の振動関数を用いることに起因し、複素数の正弦波を用いるフーリエ変換では生じない。すなわち、フーリエ変換の絶対値はシフト不変であり、多次元信号において単一方向の周波数成分を分離することができる。これはフーリエ変換で用いる式 (1) の関数の実部と虚部が、90° の位相ずれをもつヒルベルト変換対を構成しているためである。

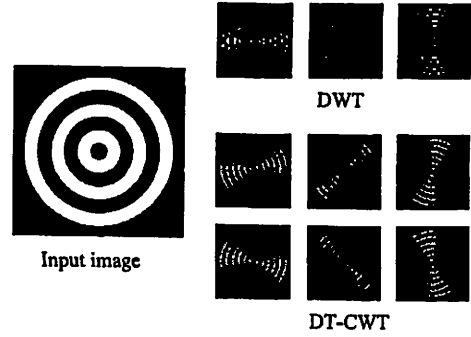


図 2 Directional selectivity of DWT and DT-CWT

$$e^{j\Omega t} = \cos(\Omega t) + j \sin(\Omega t) \quad (1)$$

この関係をウェーブレットを用いて実現したものが複素ウェーブレットである。複素ウェーブレット $\psi^c(t)$ は、2 つの実数型ウェーブレット $\psi^h(t)$, $\psi^g(t)$ を用いて式 (2) により構成される。

$$\psi^c(t) = \psi^h(t) + j\psi^g(t) \quad (2)$$

ここで、2 つの実数型ウェーブレット $\psi^h(t)$, $\psi^g(t)$ は近似的にヒルベルト変換対 $\psi^h(t) \approx \mathcal{H}\{\psi^g(t)\}$ を構成し、 $\psi^c(t)$ を用いた複素ウェーブレット変換はフーリエ変換と同様の特性をもつ。

2.2 双対木複素ウェーブレット

1 次元双対木複素ウェーブレット変換 (1D DT-CWT) は、図 1 のようにそれぞれの Tree において、実係数の最大間引きフィルタバンクを用いた DWT を実行し、その結果を複素数の実部、虚部に配置したものである。図 1 において、 $h_0(n)$, $h_1(n)$ は実部のローパス、ハイパスフィルタ、 $g_0(n)$, $g_1(n)$ は虚部のローパス、ハイパスフィルタである。ここで、フィルタバンク $\{h_0(n), h_1(n)\}$ と $\{g_0(n), g_1(n)\}$ はそれぞれ完全再構成を実現し、かつ、ヒルベルト変換対 $\psi^h(t) \approx \mathcal{H}\{\psi^g(t)\}$ が成立するために $\{h_0(n), g_0(n)\}$ は $h_0(n) \approx g_0(n - 0.5)$ が成り立つ必要がある [2]。

1D DT-CWT は、信号 $f(x) = L^2(R)$ を 1 つの複素マザーウェーブレット ψ^c と 1 つの複素スケーリング関数 ϕ^c により表現する。

$$f(x) = \sum_{l \in \mathbb{Z}} s_{j_0, l} \phi_{j_0, l}^c(x) + \sum_{j \geq j_0} \sum_{l \in \mathbb{Z}} c_{j, l} \psi_{j, l}^c(x) \quad (3)$$

ここで、 j は分解レベル、 l はシフト量、 $s_{j_0, l}$ はスケーリング係数、 $c_{j, l}$ は複素ウェーブレット係数である。

2.3 2D DT-CWT

2D DT-CWT は DWT と同様に分離型フィルタを用いる。このとき、 ψ^c , ϕ^c とその複素共役 $\bar{\psi}^c$, $\bar{\phi}^c$ の縦方向、横方向の組み合わせにより周波数上において 6 方向の帯域を分離することができる。つまり、式 (4) のように信号 $f(x, y) \in L^2(R^2)$ を 1 つの複素スケーリング関数と 6 つの複素ウェーブレット関数により表現する。

$$f(x, y) = \sum_{l \in \mathbb{Z}^2} s_{j_0, l} \phi_{j_0, l}^c(x, y) + \sum_{b \in \delta} \sum_{j \geq j_0} \sum_{l \in \mathbb{Z}^2} c_{j, l}^b \psi_{j, l}^b(x, y) \quad (4)$$

ここで、 b はサブバンドの方向をあらわし、 $\delta = \{+15^\circ, +45^\circ, +75^\circ, -15^\circ, -45^\circ, -75^\circ\}$ である。この帯域分離の様子を図2に示す。これは円形の入力画像に対するレベル2の各サブバンドを示している。DWT は3方向のエッジに分離し、そのうち対角方向のエッジは混在しているが、DT-CWT は6方向のエッジに分離できる。

3. Noise Shaping

3.1 DT-CWT における Noise Shaping

非最大間引きフィルタバンクは最大間引きフィルタバンクに比べ、サブバンド上で付加された雑音再構成信号に与える影響を小さくする特性がある。この特性を利用して、反復処理により雑音を変形し、再構成信号へ与える雑音の影響をさらに小さくする手法が Noise Shaping である [8]。

Kingsbury らはこの手法を DT-CWT へ適用し疎表現を見つける手法を提案している [4]。図3にその構成図を示す。 x は N 次元ベクトルの入力信号、 A は $M \times N$ の分解行列 ($M > N$)、 y は M 次元の DT-CWT 係数であり、 $y = Ax$ が成立する。また R は $N \times M$ の再構成行列であり、 $x = Ry$ が成立する。閾値 $f(y_i, \theta_i)$ は式 (5) のような閾値 θ を用いた非線形処理を行う。

$$\hat{y}_{2i-1} + j\hat{y}_{2i} = \begin{cases} 0 & \text{if } |y_{2i-1} + jy_{2i}| < \theta \\ y_{2i-1} + jy_{2i} & \text{otherwise} \end{cases} \quad (5)$$

2D DT-CWT では $M = 4N$ であり、再構成処理は $4N$ の係数空間から N の画像空間への射影となる。よって $4N$ の係数空間は、 N 次元の値域 S とこれに直交する $3N$ 次元の零空間 $\perp$ から構成される。値域に属するウェーブレット係数を y^S とすると $x = Ry^S$ が成立し、零空間に属するウェーブレット係数を $y^\perp$ とすると $Ry^\perp = 0$ が成立する。つまり、 y^S は再構成画像へ影響を与える係数、 $y^\perp$ は影響を与えない係数である。 $4N$ の係数空間はこれらの部分空間より式 (6) により構成される。

$$y = y^S + y^\perp \quad (6)$$

ここで、式 (5) による閾値処理を係数 y に対する雑音 d の付加としてモデル化すると、雑音 d も値域と零空間から $d = d^S + d^\perp$ により構成される。

Noise Shaping は反復処理を用いて、この雑音 d が再構成画像へ与える影響を小さくしてゆく。まず、 i 回目の閾値処理後の係数 $\hat{y}_i$ は $\hat{y}_i = y_i + d_i = y_i + d_i^S + d_i^\perp$ となり、この係数 $\hat{y}_i$ からの再構成画像 $\hat{x}$ と入力画像 x の再構成エラー e_i の分解処理により得られる更新信号 w_i は式 (7) となる。

$$w_i = kAe_i = kA(x - R\hat{y}_i) = kA(x - Ry_i - Rd_i^S) \quad (7)$$

ここで、 $Rd_i^\perp = 0$ を用いている。以上より、新たな DT-CWT 係数 y_{i+1} は式 (8) により与えられる。

$$y_{i+1} = \hat{y}_i + w_i = y_i + d_i^S + d_i^\perp + k(y_0 - ARy_i - d_i^S) \quad (8)$$

y_0 は初期の DT-CWT 係数である。式 (8) において、 $ARy_i = y_0$ より $k = 1$ の場合、

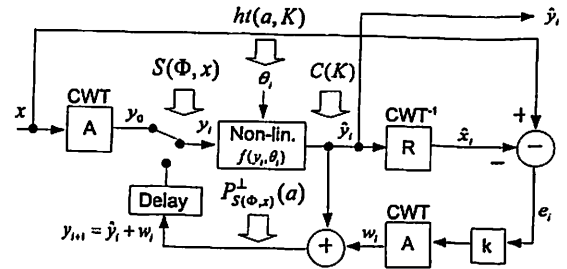


図3 Structure of Noise Shaping on DT-CWT

$$y_{i+1} = y_i + d_i^\perp \quad (9)$$

となり新たに生成される DT-CWT 係数 y_{i+1} は、 y_i の零空間 $y_i^\perp$ を雑音の零空間 $d_i^\perp$ の加算により変形することで生成される。このとき、 $Ry_{i+1} = x$ より再構成信号は不変である。つまり、Noise Shaping は零空間に属する係数 $y^\perp$ を変形する処理である。

反復処理の過程で閾値処理により生じた零空間に属する誤差 $d^\perp$ は次第に小さくなり $d^\perp \rightarrow 0$ へと近づき収束する。また、それに伴い再構成エラー e_i も $e_i \rightarrow 0$ へと近づく。反復処理の終了時、係数列 y に対する閾値処理により生じる誤差 d の大部分は値域 d^S に属する。

Noise Shaping は少数の絶対値の大きいウェーブレット係数に多くのエネルギーを閉じ込める効果があり、これらの少数の係数から再構成した信号は、処理前よりも高い Energy Compaction を与える。つまり、結果的に Noise Shaping は疎表現を近似的に求める処理に相当する。しかし、この手法で求めた疎表現の近似性能については考察がなされていない。

3.2 パラメータの定義

Noise Shaping における反復処理は2つのパラメータで定義できる。式 (5) の閾値処理は絶対値の大きな複素係数を有限個数だけ選択する操作に相当する。このとき反復処理が進むにつれ、選択閾値を徐々に増加させることで反復終了時に高い Energy Compaction を得ることができる [4]。ここで1回の反復ごとに何個の非ゼロ係数を増加するか (Gain) は、反復終了時の非ゼロ係数の総数 (NonZero) と反復回数の合計 (Iteration) の2つのパラメータを用いて式 (10) により定義できる。

$$\text{Gain} = \frac{\text{NonZero}}{\text{Iteration}} \quad (10)$$

6章において、これら2つのパラメータと求まる疎表現の関係について考察する。

4. 疎表現

4.1 問題設定と近似手法

$N \times M$ 行列 ($N < M$) Φ について、次の線形方程式を考える。

$$\Phi a = x \quad (11)$$

ここで、 $x \in R^N$ は定数、 $a \in R^M$ はこの方程式の解であり無数に存在するが、しばしば、 $\Psi = \Phi^T[\Phi\Phi^T]^{-1}$ を用いて a の

L2 ノルムを最小とする解 $a_{LS} = \Psi x$ が求められる。これに対し、疎表現は無数に存在する解の中から非ゼロ係数の個数が最少となる解を求める。つまり、疎表現は式 (12) を解くことにより与えられる。

$$\hat{a}_{0H} = \arg \min_a \|a\|_0 \quad \text{s.t.} \quad \Phi a = x \quad (12)$$

ここで、 $\|a\|_0$ は a の L0 ノルムであり非ゼロ係数の個数をあらわす。実用上はある程度の誤差を許容する条件下で疎表現を求める式 (13) を解くことが一般的である。

$$\hat{a}_{0S} = \arg \min_a \{\|a\|_0 + \lambda \|\Phi a - x\|_2^2\} \quad (13)$$

ここで λ はラグランジュ乗数である。ただし、これらの問題は NP 困難であり計算を行うことが非常に複雑である。そこで、これらの問題を近似的に解く手法が提案されており、Matching Pursuit (MP) [5] と Basis Pursuit (BP) または Basis Pursuit De-Noising (BPDN) [6] が良く知られている。MP は逐次的に対象物を最もよく表現する基底を選び出す手法であり、BP (または BPDN) は L0 ノルムを L1 ノルムへ置き換え線形計画法を用いて疎表現を求める手法である。一方で、Mancera らは、L0 ノルムを直接用いて疎表現を求める手法を提案している [7]。

4.2 L0 ノルムベースの疎表現近似

Mancera らは、 λ を用いて式 (13) を解くことが定数 K を用いて $\|\hat{a}_{0S}\|_0 = K$ という条件下で式 (13) を解くことと等価であることを利用し、式 (14) のように問題を設定している。

$$\begin{aligned} \hat{a}(K) &= \arg \min_{a \in C(K)} \|\Phi a - x\|_2 \\ \text{s.t.} \quad C(K) &= \{a \in \mathbb{R}^M : \|a\|_0 = K\} \end{aligned} \quad (14)$$

これは x とのユークリッド距離を最小化する K 個の非ゼロ係数を持つ a を探していると解釈することができる。 $\hat{a}(K)$ を求めるため、式 (11) の解のアフィン集合 $S(\Phi, x)$ と $C(K)$ の間の交互直交射影を繰り返す [10]。 $C(K)$ から $S(\Phi, x)$ への直交射影は $P_{S(\Phi, x)}^\perp = a + a_{LS} - \Psi \Phi a$ 、 $S(\Phi, x)$ から $C(K)$ への直交射影は $ht(a, K)$ により与えられる。ここで、 $ht(a, K)$ は $C(K)$ へのユークリッド距離を最小にする射影であり、最も大きい値から K 個の係数を選択する操作に相当する。 $K = 1$ における反復処理の幾何的な遷移を図 4 に示す。 Mancera らは、この手法を Curvelet に適用し BP (または BPDN) よりも高い Energy Compaction が得られることを示している。

5. 提案手法

5.1 Noise Shaping の疎表現解釈

3 章で述べた Noise Shaping と 4.2 節で述べた L0 ノルムベースの疎表現近似手法は Noise Shaping のパラメータ $k = 1$ の場合等価である。まず、式 (5) における閾値処理は $ht(a, K)$ に相当し、最大値から K 個の非ゼロ係数を選択していると考えることができる。また、 $P_{S(\Phi, x)}^\perp$ は式 (15) のように変形することができる。

$$\begin{aligned} P_{S(\Phi, x)}^\perp &= a + a_{LS} - \Psi \Phi a \\ &= a + \Psi(x - \Phi a) \end{aligned} \quad (15)$$

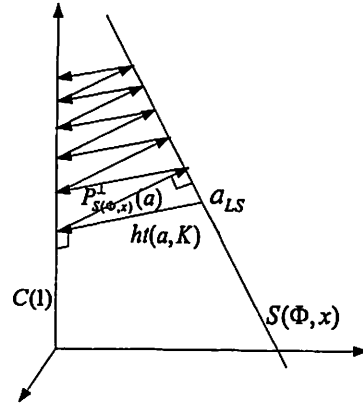


図 4 L0-norm minimization through alternate projections

これは Noise Shaping において、更新信号 w_i を $\hat{y}_i$ に加算する操作に相当する。これらの対応関係を図 3 に示す。この等価性より、疎表現を求めることは少なくとも零空間の係数を変形し、有限個の非ゼロ係数から再構成した信号の誤差が値域から発生するような状態を作り出すことであると考えられる。

Mancera らの提案手法は K が固定であるが、Noise Shaping との等価性より K を反復処理とともに増加させることで、彼らの手法は改善されることがわかる。これは Matching Pursuit のように、近似精度の高い基底から順に選択することで近似精度の悪い基底が選択されにくくなるためである。ただし、Noise Shaping は一度選択した基底は終了時まで保持されるという特性があるため、Huggings らが提案しているような不適切になった基底を除去する操作を加えることで、より理想的な疎表現に近づく可能性がある。

L0 ノルムベースの疎表現近似手法との等価性より、複素ウェーブレットにおける Noise Shaping を用いた疎表現は Curvelet の場合と同様に Basis Pursuit (または BPDN) よりも理想的な疎表現に近いものを与えている可能性がある。しかし、近似手法の性能は用いる基底によって変わるため、Basis Pursuit (または BPDN) との厳密な比較は複素ウェーブレットの行列 Φ の Mutual Coherence [9]などを考慮する必要がある今後の課題である。

5.2 提案手法

3 章の Noise Shaping を用いた複素ウェーブレットの疎表現では、式 (5) の閾値処理において複素係数単位で最大値より係数を選択する。しかし、L0 ノルムベースの $ht(a, K)$ は最大値より K 個の非ゼロ係数を選択する操作であり、必ずしも複素係数単位で係数を選択する必要はない。そこで、複素数単位で選択するという制約を排除し、式 (16) のように実部・虚部からそれぞれ独立に係数を選択する手法を提案する。

$$\hat{y}_{2l-1} + j\hat{y}_{2l} = \begin{cases} y_{2l-1} + jy_{2l} & \text{if } |y_{2l-1}| > \theta \text{ and } |y_{2l}| > \theta \\ y_{2l-1} & \text{if } |y_{2l-1}| > \theta \\ jy_{2l} & \text{if } |y_{2l}| > \theta \\ 0 & \text{otherwise} \end{cases} \quad (16)$$



(a) DWT(31.74dB) (b) conv.(31.94dB) (c) Prop.(34.11dB)

図5 Reconstructed Image

この手法は次章の実験結果で示すように高い Energy Compaction が得られる。これは、式 (5) によって選択される複素係数の実部と虚部の両基底が必ずしも近似精度の良い基底であるとは限らないからである。これに対し、式 (16) は実部と虚部の基底を独立に考えることで、近似精度の良い基底のみが選択される。しかし、式 (16) により選ばれた係数は複素係数ではなく、シフト不変性の特性は失われる。また、従来手法と提案手法の係数選択方法は統合することが可能である。たとえば、エッジ近辺の係数はシフト不変性の成立する複素数として式 (5) を用いて選択し、それ以外の領域の係数は実部、虚部として式 (16) を用いて選択することができる。このときの Energy Compaction は両手法の間に位置すると予想される。

6. 実験結果と考察

本実験では、DT-CWT としてレベル 1 において (13,19)-tap odd filter と (12,16)-tap even filter, レベル 2 以降において 18-tap Q-shift filter を使用し [1], DWT として Daubechies-9/7 filter を使用する。また、レベル 5 まで分解処理を行い、実験画像として Lena (256×256: グレースケース) を用いる。

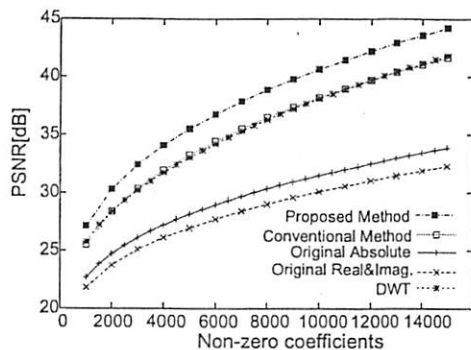
6.1 係数選択方法と Energy Compaction

Noise Shaping において、複素数として係数を選択する場合 (conventional) と実部、虚部から独立に係数を選択する場合 (proposed) の Energy Compaction を図 6(a) に示す。ここでの疎表現は PSNR (Peak Signal to Noise Ratio) が最も高くなる $k=1.6$ とし、各反復において 1 個ずつ非ゼロ係数を選択 ($Gain=1$) することで求めている。PSNR は反復処理の終了個数 (NonZero) と同数の非ゼロ係数から再構成した画像を用いて計算した。この場合、本手法において最も良い疎表現が求められるが、膨大な計算処理時間を要する。

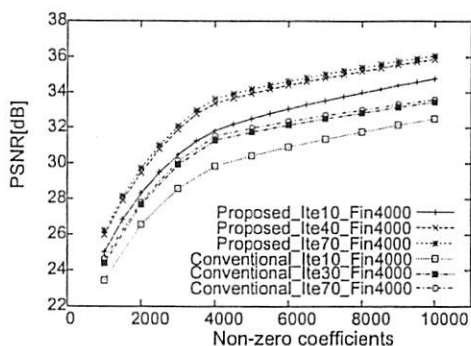
図 6(a) より、従来手法の Energy Compaction は DWT とほぼ同じであるのに対し、提案手法の Energy Compaction はこれより約 1.5~2.5[dB] 高いことがわかる。また、4000 個の非ゼロ係数を用いて再構成した画像の一部を図 6 に示す。従来手法と DWT の PSNR はほぼ同じであるが複素ウェーブレットを用いる方がエッジ近辺の再構成精度が良く主観的に高い品質である。

6.2 パラメータと Energy Compaction

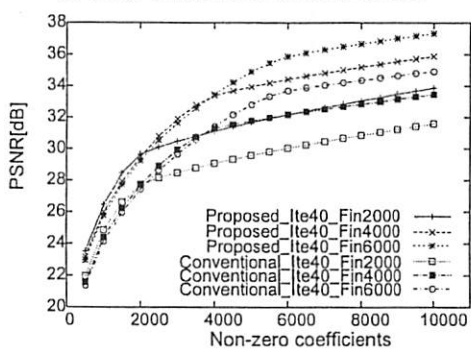
本節では、3.2 節の各パラメータと求まる疎表現の Energy Compaction の違いについて述べる。まず、終了個数を固定 ($NonZero=Const.$) し反復回数を変化させて求めた疎表現の Energy Compaction を図 6(b) に示す。反復回数を増加させることは、式 (10) より 1 回の反復における非ゼロ係数の増加個



(a) Energy Compaction of Noise Shaping



(b) Energy Compaction by number of Iteration



(c) Energy Compaction by number of NonZero

図6 Energy Compaction of Sparse Representation

数を少なくすることに相当し、不適切な基底を選ぶ可能性が低減されるため高い Energy Compaction を得ることができる。

次に、反復回数を固定 ($Iteration=Const.$) し終了個数を変化させて求めた疎表現の Energy Compaction を図 6(c) に示す。この場合、終了個数を境にして非ゼロ係数の個数に対する PSNR 向上率が異なることがわかる。これは終了個数に達するまでの非ゼロ係数で疎表現を求めるため、それ以降の係数において対象物を疎に表現できていないためである。

6.3 パラメータと相互情報量

本節では、3.2 節の各パラメータと求まる疎表現の相互情報量の違いについて述べる。相互情報量は 2 変数間の依存関係を与える基準として利用され、ウェーブレット係数に適用することが可能である。ここで、対象とするウェーブレット係数を X , parent (同空間位置の 1 レベル低域の係数) を PX , neighbors

表 1 Mutal Information of DT-CWT (number of Iteration)

Iteration	original			10			40			70		
Mutal Info.	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$
Conventional	0.44	0.52	0.88	0.40	0.46	0.83	0.35	0.46	0.80	0.31	0.46	0.79
Proposed	0.09	0.09	0.21	0.08	0.09	0.17	0.07	0.09	0.16	0.07	0.09	0.15

表 2 Mutal Information of DWT

Mutal Info.	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$
horizontal	0.12	0.11	0.25
diagonal	0.09	0.09	0.18
vertical	0.19	0.14	0.38

表 3 Mutal Information of DT-CWT (number of NonZero)

NonZero	2000			4000			6000		
Mutal Info.	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$	$I(X;PX)$	$I(X;CX)$	$I(X;NX)$
Conventional	0.40	0.48	0.86	0.35	0.46	0.80	0.28	0.45	0.74
Proposed	0.08	0.09	0.18	0.07	0.09	0.16	0.07	0.09	0.14

(同一サブバンド内の 8 近傍) を NX , cousins (同空間位置, 同レベルの他方向サブバンド) を CX とし, X と各変数間の相互情報量を $I(X;PX)$, $I(X;NX)$ および $I(X;CX)$ と定義し, 文献 [12] に基づき推定する. このとき, $I(X;PX)$, $I(X;NX)$ および $I(X;CX)$ はそれぞれレベル間, サブバンド内およびサブバンド間の相関性をあらわしている. 表 1, 3 は DT-CWT のレベル 1 の 15° 方向サブバンドに属する X に着目した相互情報量, 表 2 は DWT のレベル 1 の各方向サブバンドに属する X に着目した相互情報量をあらわしている.

表 1 において反復回数が多いほど, つまり理想的な疎表現に近づくほど相互情報量が小さくなる. よって, 疎表現はウェーブレット係数間の相関性を小さくする. また, 表 3 において終了個数が多いほど相互情報量が小さくなっている. これは終了個数が少ないときは, より低域成分をあらわす基底で疎表現を見つけようとするため, 高域成分の基底に関連する係数が疎表現の影響を受けにくいためと考えられる.

6.4 符号化への考察

Noise Shaping を符号化へ適用する場合, 2 つのパラメータは次のような基準で決定できる. 図 6(c) より Noise Shaping で求めた疎表現は, 終了個数 (NonZero) 近辺の非ゼロ係数では高い PSNR を与えるが, それ以外の非ゼロ係数の個数については PSNR が低くなる. よって, 送信する非ゼロ係数の個数と終了個数を一致させると, その個数において最も高い PSNR が得られる. この個数はターゲットビットレートにより変化するため, ビットレートより推定する必要がある. また, 表 3 より終了個数が少ないほど反復処理終了後も多くの冗長性が残り, 付加的な処理を行うことでさらに符号化効率を高めることができる. 反復回数 (Iteration) は, エンコードにおける処理時間を考慮し決定する. このとき, 図 6(c), 表 1 より反復回数が多いほど高い Energy Compaction が得られるが, ウェーブレット係数間の相関性が低くなる.

複素数として係数を選択した場合は実部, 虚部から独立に選択した場合と比べ Energy Compaction は低い, 反復処理終了時においてもウェーブレット係数間に非常に高い相関が残っている. また, 絶対値と位相に分けて量子化, エントロピー符号化を行うことが可能である [14].

7. まとめと今後の課題

本稿では, 複素ウェーブレットにおいて疎表現を求める Noise Shaping が L0 ノルムベースの疎表現近似手法と等価であることを示し, その等価性に基づき, 従来とは異なる係数選択

の手法を提案した. 実験より, 提案手法において高い Energy Compaction を得られることを確認した. また Noise Shaping の反復処理を定義する 2 つのパラメータと求める疎表現の関係を Energy Compaction と相互情報量を用いて考察した.

今後は複素ウェーブレットを用いた画像符号化の実現を目指し, Noise Shaping の性能を Basis Pursuit (または BPDN) と比較すると共に, 求めた疎表現に残されている冗長性をさらに削減する手法を検討する.

文 献

- [1] N. G. Kingsbury, "Complex wavelet for shift invariant analysis and filtering of signals," ACHA, vol. 10, no. 3, pp. 234-253, May 2002.
- [2] I. W. Selesnick, R. G. Baraniuk and N. G. Kingsbury, "The Dual-Tree Complex Wavelet Transform," IEEE Signal Processing Magazine, Nov. 2005.
- [3] R. M. Figueras i Ventura, et al., "Low-rate and flexible image coding with redundant representations," IEEE Trans. Image Processing, Vol. 15, No. 3, Mar. 2006.
- [4] N. G. Kingsbury and T. H. Reeves, "Iterative image coding with overcomplete complex wavelet transforms," Proc. Conf. on VCIP, July 2003.
- [5] S. G. Mallat and Z. Zhang, "Matching Pursuit With Time-Frequency Dictionaries," IEEE Transaction on Signal Processing, vol. 41, no. 12, pp. 3397-3415, December 1993.
- [6] S. Chen, D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM J. Scientific Comp., vol. 20, pp. 33-61, 1999.
- [7] L. Mancera and J. Portilla, "L0-norm-based sparse representation through alternate projections," IEEE ICIP, Oct. 2006.
- [8] H. Bölcskei, F. Hlawatsch, "Oversampled filter banks: Optimal noise shaping, design freedom, and noise analysis," Proc. IEEE ICASSP, vol. 3, pp. 2453-2456, April 1997.
- [9] D. L. Donoho, M. Elad and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Information Theory, vol. 52, pp. 6-18, Jan. 2006.
- [10] D. C. Youla, "Generalized Image Restoration by the Method of Alternating Orthogonal Projections," IEEE Trans. Circuits and Syst., CAS-25, 9, 1978.
- [11] P. S. Huggins and S. W. Zucker, "Greedy Basis Pursuit," IEEE Trans. Signal Processing, vol. 2, pp. 3760-3772, July 2007.
- [12] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," IEEE Trans. Image Processing, vol. 10, no. 11, pp. 1647-1658, 2001.
- [13] J. Yang, W. Xu Q. Dai and Y. Wang, "Image Compression using 2D Dual-Tree Discrete Wavelet Transform (DDWT)," IEEE ISCAS, May 2007.
- [14] T. H. Reeves and N. G. Kingsbury, "R-D quantisation of complex coefficients in zerotree coding," IEEE SSP, 2001.

複素ウェーブレットの疎表現における係数選択手法の検討

A Study for Coefficients Selection Method on Sparse Representation of Complex Wavelet

高橋 良知†

Yoshitomo TAKAHASHI

渡辺 裕†

Hiroshi WATANABE

1. まえがき

離散ウェーブレット変換 (DWT) は方向分離性の乏しい変換である。一方で、複素ウェーブレット変換 (CWT) は高い方向分離性をもつ変換である。方向分離性をもつ変換は、エッジなどの曲線を方向性基底により近似するため、画像構造に基づく近似が可能である。しかし、この変換は変換後にデータ数が増加する冗長変換であることが一般的である。これは冗長性の削減という画像符号化の目的に反するため、画像符号化への応用はほとんどなされていない。冗長変換では疎表現と呼ばれる係数列が存在する。疎表現係数列は信号を最少数の非ゼロ係数で表現したものであり、画像符号化に有効な係数列である。

我々は、方向分離性と冗長性をもつ CWT の疎表現を用いた画像符号化の検討を行っている。本稿では、疎表現の近似解法である l^0 ノルムベース交互射影法に基づき、Noise Shaping における新しい係数選択手法を提案する。また、非ゼロ係数の個数に対する PSNR と全符号量により性能を評価する。

2. 複素ウェーブレット変換

CWT は 2 つの実数ウェーブレットを実部と虚部の基底とした複素数変換である [1]。これはフーリエ変換における正弦波をウェーブレットに置き換えた変換といえる。CWT はシフト不変性が成立し、2 次元 CWT は $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ の 6 方向の画像エッジを異なるサブバンドに分離する。また、CWT は 2 つの最大間引きフィルタバンクを並列に配置した並列木構造により実装される。このため、2 次元の入力信号に対して 4 倍のデータ数が発生する冗長変換となる。

3. 冗長変換の疎表現

3.1 疎表現問題

n 次元ベクトル $y \in \mathbb{R}^n$ を辞書 $\mathcal{D} = \{\phi_i \in \mathbb{R}^n, i = 1, \dots, m\}$ の線形結合により表現する問題を考える。

$$y = \sum_i \alpha_i \phi_i \quad (1)$$

冗長変換では $m > n$ であり、辞書 $\mathcal{D}$ は過完備な基底から構成されている。このとき、式 (1) の解 $\alpha \in \mathbb{R}^m$ は無数に存在する。このうち疎表現とは最少数の基底 (非ゼロ係数) を用いて信号 y を表現した解である。これは式 (2) の最適化問題の解として求められる。

$$\text{Minimize } \|\alpha\|_0 + \lambda \|\mathbf{D}\alpha - y\|_2^2 \quad (2)$$

式 (2) は誤差を含む場合の疎表現問題である。ここで、 $\mathbf{D}$ は $n \times m$ 行列、 $\|\alpha\|_0$ は α の l^0 ノルムであり、非ゼロ係数の個数である。式 (2) は NP 困難のため、マッチング追跡や基底追跡などの近似解法が必要である。

3.2 l^0 ノルムベース交互射影法

本稿では、近似解法として l^0 ノルムベース交互射影法を用いる。この手法では式 (2) の代わりに式 (3) の問題より近似解を求める。

$$\begin{aligned} \hat{\alpha}(K) &= \arg \min_{\alpha \in C(K)} \|\mathbf{D}\alpha - y\|_2 \\ \text{s. t. } C(K) &= \{\alpha \in \mathbb{R}^m : \|\alpha\|_0 = K\} \end{aligned} \quad (3)$$

解 $\hat{\alpha}(K)$ を求めるため、式 (1) の解集合 $S(\mathbf{D}, y)$ と $C(K)$ の間の交互射影を繰り返す [2]。ここで、 $S(\mathbf{D}, y)$ は式 (1) のアフィン集合で凸集合であるが、 $C(K)$ は凸集合ではない。そのため、収束解は局所解である。 $C(K)$ から凸集合 $S(\mathbf{D}, y)$ への射影は直交射影 $P_{S(\mathbf{D}, y)}^\perp$ が存在する。凸集合 $S(\mathbf{D}, y)$ から $C(K)$ への射影 $ht(\alpha, K)$ は $C(K)$ と $S(\mathbf{D}, y)$ のユークリッド距離を最小にするために最大値から K 個の係数を選択する。

3.3 CWT における Noise Shaping

Kingsbury らは CWT に対して、雑音変形手法である Noise Shaping を適用し疎表現を求めている [3]。Noise Shaping は反復処理に基づく手法であり、反復処理の一部で式 (4) の閾値処理がおこなわれる。

$$\hat{y}_r + j\hat{y}_i = \begin{cases} 0 & \text{if } |y_r + jy_i| < \theta \\ y_r + jy_i & \text{otherwise} \end{cases} \quad (4)$$

ここで、 y_r , y_i は複素数の実部、虚部である。 $\hat{y}_r$, $\hat{y}_i$ は閾値処理後の実部、虚部である。

†早稲田大学大学院 国際情報通信研究科
Graduate School of Global Information and Telecommunication Studies, Waseda University.

4. 提案手法

我々は、CWT における Noise Shaping が l^0 ノルムベース交互射影法と等価な手法であることを示した [4]。等価性より、Noise Shaping における式 (4) の閾値処理は l^0 ノルムベース交互射影法における $ht(\alpha, K)$ に相当する。 $ht(\alpha, K)$ は最大値から K 個の係数を選択する処理であるため、式 (4) の閾値処理は複素数単位で選択する必要はない。そこで、我々は実部と虚部に対して、独立に閾値処理を適用する手法を提案する。

$$\hat{y}_r = \begin{cases} 0 & \text{if } |y_r| < \theta \\ y_r & \text{otherwise} \end{cases} \quad \hat{y}_i = \begin{cases} 0 & \text{if } |y_i| < \theta \\ y_i & \text{otherwise} \end{cases} \quad (5)$$

Noise Shaping はマッチング追跡と同様の貪欲アルゴリズムである。そのため、基底の選択方法が重要となる。提案手法では、実部と虚部の基底を独立に選択することで、複素数単位で選択する場合と比較し、誤基底を選択する可能性を低減している。

また、疎表現の性能は近似解法と辞書の定義により決まる。方向性基底の複素ウェーブレットを $\psi_j^c = \psi_j^r + j\psi_j^i$ とすると、辞書 $\mathcal{D}$ は、従来手法では $\mathcal{D}_c = \{\psi_j^c \in \mathbb{R}^n, j = 1, \dots, 2n\}$ 、提案手法では $\mathcal{D}_p = \{\psi_j^r, \psi_j^i \in \mathbb{R}^n, j = 1, \dots, 2n\}$ となる。 ψ_j^r と ψ_j^i はヒルベルト変換対であるためシフトの関係にある。このため、提案手法ではより密な方向性基底の集合を辞書 $\mathcal{D}_p$ として定義していることになる。

5. 実験結果

Barbara (512×512) において、非ゼロ係数の個数に対する PSNR を図 1 に示す。提案手法では、DWT や従来手法と比較し、同数の非ゼロ係数で約 1.0–3.0 [dB] 高い PSNR が得られている。これより、提案手法は従来手法よりも疎表現の近似解精度が高い。また、全符号量における比較を図 2 に示す。CWT では各レベルのエントロピーの和に基づき全符号量を求めた。量子化として、提案手法にはスカラ量子化、従来手法にはスカラ量子化または複素数のベクトル量子化を用いた。DWT は算術符号化なしの SPIHT 符号化を行った [5]。図 2 より、提案手法は SPIHT 符号化より約 0.2–0.5 [dB] 高い符号化効率となる。また、提案手法は従来手法のスカラ量子化より高い符号化効率となるが、従来手法のベクトル量子化よりは性能が劣る。これより、提案手法の疎表現においても適したベクトル量子化を考慮する必要があり、今後の課題である。

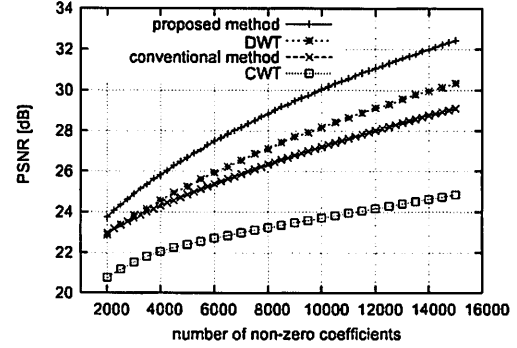


図 1: 非ゼロ係数へのエネルギー集約度

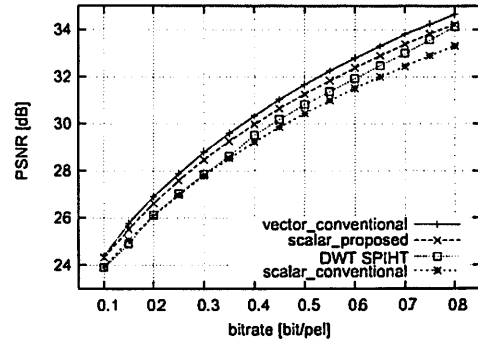


図 2: 全符号量

6. まとめ

本稿では、CWT における疎表現近似手法において新しい係数選択手法を提案した。非ゼロ係数へのエネルギー集約度と全符号量による実験より、その有効性を確認した。

参考文献

- [1] I.W. Selesnick, *et al.*, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, Nov. 2005.
- [2] L. Mancera, *et al.*, "L0-norm-based sparse representation through alternate projections," *IEEE ICIP*, Oct. 2006.
- [3] N. G. Kingsbury *et al.*, "Iterative image coding with overcomplete complex transform," *IEEE VCIP*, July 2003.
- [4] 高橋, 渡辺, "双対木複素ウェーブレットにおける疎表現の検討," *AVM58-06*, Sep. 2007.
- [5] A. Said, *et al.*, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. CirSys Video*, June 1996.

複素ウェーブレット変換の疎表現における係数予測手法の検討

A Study for Coefficients Prediction on Sparse Representation of Complex Wavelet Transform

高橋 良知 Yoshitomo TAKAHASHI 渡辺 裕 Hiroshi WATANABE

早稲田大学大学院 国際情報通信研究科
Graduate School of Global Information and Telecommunication Studies, Waseda University.

1 はじめに

離散ウェーブレット変換 (DWT) は変換係数から得られる画像の方向情報が乏しい。一方で、複素ウェーブレット変換 (CWT) は絶対値においてシフト不変性が成立し、2次元 CWT は $\pm 15^\circ$, $\pm 45^\circ$, $\pm 75^\circ$ の6方向の画像エッジを分離する変換である [1]。また、2次元 CWT は変換後のデータ数が増加する冗長変換でもある。冗長変換では疎表現と呼ばれる係数列が存在する。CWT の疎表現係数列による画像符号化は、全符号量において DWT の SPIHT 符号化より約 0.5-1.0[dB] 高い符号化効率を得られている [3]。

本稿では、CWT の疎表現係数列において、係数の絶対値の冗長性を利用し、高域係数の絶対値を低域係数の絶対値の補間拡大処理より予測する手法を検討する。

2 ウェーブレット係数予測の従来手法

本手法は低域係数より高域係数を予測するという点で、ウェーブレット空間における画像の高解像度化手法と類似している。高解像度化手法では低解像度画像のウェーブレット係数列を対象とし、係数の統計情報より高解像度画像の最高周波数成分に対応するサブバンド係数を予測する [4]。一方で、本研究の対象は疎表現係数列である。疎表現係数列は少数の非ゼロ係数と多数のゼロ係数からなる疎な係数列であり、係数の統計情報も通常のウェーブレット係数とは異なる。また、本研究では、予測対象サブバンドが最高周波数サブバンドのみでないため、従来の予測手法を適用することは難しい。

3 CWT 係数の絶対値予測手法

ステップエッジに対する CWT 係数の絶対値はガウシアン分布に近似され、その分散は低域サブバンドほど大きくなる。よって、CWT 係数のサブバンドの絶対値はガウシアンピラミッド構造に近似される。この特性より、低域のサブバンドの絶対値を補間拡大したものを高域の予測サブバンドの絶対値とする手法を提案する。提案手法の概要を図1に示す。各レベル、各方向のサブバンドにおいて、図1の処理を低域側から独立に行う。まず、低域のサブバンド $S_{m,n}^{j+1}$ を零補間とローパスフィルタを用いて拡大する。次に、同一のレベル・方向で固定

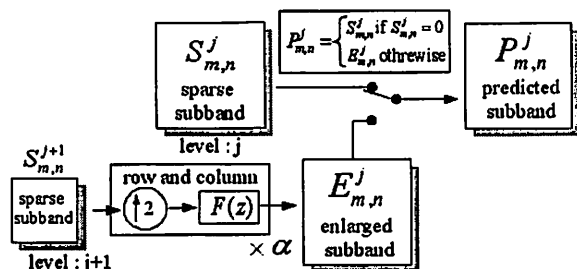


図1 The flow of coefficients prediction

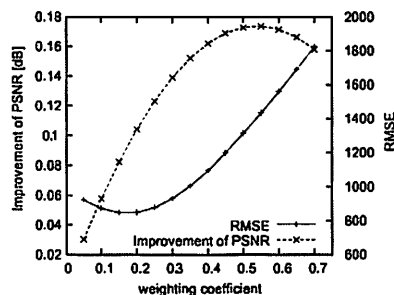


図2 PSNR and RMSE in level 4, -75°

値の重み係数 α を用いてエネルギーを変更し、拡大サブバンド $E_{m,n}^j$ を得る。レベル j のサブバンド $S_{m,n}^j$ がゼロ係数である場合、拡大サブバンド $E_{m,n}^j$ の係数値で置き換え、予測サブバンド $P_{m,n}^j$ を生成する。

4 実験結果

重み係数 α を変化させたときの、CWT 係数の真値と予測値の RMSE と予測値からの再構成画像の PSNR 向上値を図2に示す。冗長変換は直交変換と異なり変換係数間に依存性が存在する。このため、ウェーブレット係数における最小誤差が再構成画像で最大の PSNR を与えないという特性がある。

また、適切な重み係数 α を用いてレベル4とレベル3のサブバンドを予測した結果を表1に示す。表1は、真値を用いたときの PSNR 向上に対する提案手法による PSNR 向上の割合であり、達成率である。予測した絶対値の位相には符号情報に相当する 2[bit] を与えている。予測を行うことで 35~50[%] 程度の達成率が得られており、本予測手法は有効であるといえる。

5 まとめと今後の課題

本稿では、CWT の疎表現係数列の絶対値の冗長性を利用する一手法として、低域サブバンドから高域サブバンドを予測する手法の検討を行った。今後は、CWT の疎表現における位相係数を予測する手法を検討し、CWT 画像符号化の符号化効率向上を目指す。

参考文献

- [1] I.W. Selesnick, et al., "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, Nov. 2005.
- [2] L. Mancera, et al., "L0-norm-based sparse representation through alternate projections," *IEEE ICIP*, Oct. 2006.
- [3] 高橋, 渡辺, "複素ウェーブレットの疎表現における係数選択手法の検討," 情処全大, Mar. 2008.
- [4] D. H. Woo, et al., "Image interpolation based on inter-scale dependency in wavelet domain," *IEEE ICIP*, Oct. 2004.

表1 The achievement rate by prediction method

predicted level	direction of subband						All
	15°	45°	75°	-15°	-45°	-75°	
from 5 to 4 (%)	32.9	27.3	48.6	37.1	39.4	40.2	34.7
from 4 to 3 (%)	50.0	50.0	57.1	56.7	59.2	54.4	47.8