

平成18年度 修士論文

Research on Time Series Analysis  
for Spatio-Temporal Image Coding  
時空間画像符号化のための時系列解析に関する研究

指導教授 渡辺 裕 教授

2007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

4205A117-2

加 藤 幸 一

Koichi KATO

## 目次

<b>第 1 章</b>	<b>序論</b>	<b>1</b>
1.1	研究の背景	1
1.2	研究の目的	2
1.3	本論文の構成	3
<b>第 2 章</b>	<b>関連符号化技術</b>	<b>4</b>
2.1	はじめに	4
2.2	JPEG 2000	4
2.2.1	概要	4
2.2.2	符号化アルゴリズム	6
2.2.3	動向	12
2.3	Motion-JPEG 2000	13
2.3.1	特徴	13
2.3.2	デジタルシネマ	14
2.4	H.264/AVC	15
2.4.1	概要	15
2.4.2	符号化アルゴリズム	16
2.4.3	プロファイル規定	17
2.4.4	アプリケーション	19
2.5	解像度変換	19
2.5.1	間引き	19
2.5.2	デシメータ	19
2.5.3	MCTF	19
2.6	終わりに	20
<b>第 3 章</b>	<b>時系列信号</b>	<b>21</b>
3.1	はじめに	21
3.2	時空間画像	21
3.2.1	概要	21
3.2.2	関連研究	21
3.3	時空間画像の符号化	23
3.3.1	提案手法	23

3.3.2	問題点	25
3.4	高フレームレート映像	26
3.4.1	概要	26
3.4.2	問題点	27
3.4.3	時間解像度変換	27
3.5	おわりに	27
<b>第4章</b>	<b>時空間画像符号化の改善手法</b>	<b>28</b>
4.1	はじめに	28
4.2	実験と考察	28
4.2.1	PSNRの揺らぎについて	28
4.2.2	SNRの急激な低下について	33
4.2.3	シーケンスへの依存度について	34
4.3	フレーム間予測	36
4.4	おわりに	40
<b>第5章</b>	<b>高フレームレート映像の時間解像度変換</b>	<b>42</b>
5.1	はじめに	42
5.2	高フレームレート映像	42
5.3	符号化効率	44
5.4	時間解像度変換	46
5.5	おわりに	49
<b>第6章</b>	<b>結論</b>	<b>50</b>
6.1	総括	50
6.2	今後の課題	51
	謝辞	52
	参考文献	53
	図一覧	56
	表一覧	57
	研究業績	58

# 第1章

## 序論

### 1.1 研究の背景

近年、デジタル技術の発達による情報通信分野の変化は著しく、とりわけ映像コンテンツの流通はとどまることを知らない。テレビ放送のデジタル化やネットワークの広帯域化、またはそれに伴うサービスの多様化により、今後ますますあらゆるコンテンツが、様々なメディアを通じてユーザの目に行き届くことが想像される。映像コンテンツの起点から終点までがすべてデジタル化され、ユーザはその恩恵を享受し、任意に映像を配信から視聴まで行うことが可能となる<sup>(1)</sup>。

流通環境の多様さと同じくして、コンテンツの性質も様々である。非常にハイスペックな機材で大規模な予算で制作されたものから、家庭で制作されたものまで同等に行き交う時代が来るであろう。視聴者はその中から自分の希望するコンテンツを選択するようになる。視聴される映像コンテンツもそれに見合うだけの品質を伴っていることが期待される。撮影に使用されるカメラの種類もバラエティに富んでおり、デジタル技術の発展とともに、高品質なものが安価で入手できるようになり、家庭でも相当な質のコンテンツが制作可能になった<sup>(2)</sup>。また、プロユースのものでは、Full HD 解像度まで対応したものや、100,000 フレーム/秒以上撮影可能なものまで、用途に応じた選択が可能となっている。そうしたコンテンツが視聴される環境も多彩で、外出先における携帯電話などの QVGA (320x240) 程度のものからホームシアターにおける Full HD (1920x1080)、さらには劇場スクリーンでの 4K (4096x2160) など、無数の解像度に対する需要がある。NHK によってスーパーハイビジョン (7680x4320) という次世代の超高精細映像システムの研究もなされており、今後のさらなる高画質化が期待される<sup>(3)</sup>。

表示方式も、古くはブラウン管に始まり、LCD (Liquid Crystal Display) や PDP (Plasma Display Panel)、SED (Surface-conduction Electron-emitter Display) など枚挙に暇がない。いずれも解像度の向上とともに、応答速度、コントラスト比なども改善され、より高精細な映像を表示できるよう進化を続けている<sup>(4)</sup>。

しかし、いかなる流通システムや表示デバイス、テレビや携帯電話、PC モニタといった視聴スタイルであっても、より高画質な映像を見たいという欲求は普遍なものであろう。ユーザ側は、自分の環境における最大限の映像を体感したいはずである。そして、その環境に合

わせて映像コンテンツが形態を変える必要がある。複雑な処理・操作を伴わず、よりシームレスに体感されることが望ましい。

ネットワークの広帯域化、蓄積装置の大容量化は今後も果てしなく進んでいくことが予想されるが、データ容量もそれに合わせて拡大していくものであり、決して必要十分なスペックまで辿り着くことはない。

そのため、動画像データの符号化によって品質を保ちながら、符号量を削減する検討が数多くなされ、一定の成果に達している。動画像符号化の国際標準化を担う機関として MPEG (Moving Picture Experts Group) と ITU-T (International Telecommunication Union-Telecommunication Standardization Sector) がある。近年、本格サービスが開始された地上デジタル放送で採用された MPEG-2 や、MPEG と ITU-T が共同で標準化を行い、次世代ディスク Blu-ray や HD DVD で採用されている H.264/MPEG-4 AVC が今後も主流として広く利用されるであろう。また、動き補償と DCT (Discrete Cosine Transform) を要素技術とした MPEG-2, MPEG-4 とは対照に、デジタルシネマなどの高精細な表現が求められるシーンにおいて、DWT (Discrete Wavelet Transform) を用いた JPEG 2000 符号化も地位を確立していくであろう。1つのソースから流通の形態・端末の機能に合わせて必要なデータを選択可能にした符号化を行うスケーラビリティ機能についても検討されている。

映像デバイス全般の高性能化に伴い、映像メディアの高精細化が進む中で、空間解像度と同様に時間解像度の向上が期待される。しかし、時間解像度に関しては、現在用いられているフレームレートは人の視覚系に必ずしも適合せず、検討も不十分である<sup>(5)</sup>。また、高フレームレート化による符号量の増加も問題となり、そうした映像に対する効率的な符号化手法が求められる<sup>(6)</sup>。

## 1.2 研究の目的

本研究では、時空間画像を用いた時系列信号の符号化について検討を行う。動画像の符号化で広く用いられている MPEG では、フレームレート 30 [fps]、フィールドレート 60 [fps] を想定し、これが広く普及している。これ以下のフレームレートへの変換は、フレームの間引きによって実現されるが、シャノンの標本化定理に適っているとはいえない。また、SVC (Scalable Video Coding) では MCTF (Motion Compensated Temporal Filtering) による手法が採られているが、想定された以上のフレームレートに関しては議論の余地がある。

また、メモリの大容量化により、一度に多くのフレームを扱うことが可能となり、これまでと同じ実時間上の長さでも、より時間解像度の高い信号を1つの単位として扱うことが出来る。そのような信号に、MPEG で行われるフレーム間予測を適用した場合、フレームレートの上昇とともに、動きベクトル数の増加により符号量の爆発的な増加が懸念され、検討が必要である。

そこで、本研究では空間軸と時間軸からなる画像を時空間画像と定義し、空間画像と同等な処理によって時空間画像を符号化し、動画像における時系列信号の符号化に対する有効性について検討する。また、時系列信号を階層的に符号化することで、時間解像度の変換を実

現し、適応的に所定のフレームレートにおける、高品質な動画像の提供の実現について検討を行う。

### 1.3 本論文の構成

以下に本論文における本章以降の構成を示す。

第2章 本研究で用いられる要素技術について述べる。まず、本研究の基盤となる JPEG 2000 符号化技術について説明する。次に、JPEG 2000 の拡張フォーマットである Motion-JPEG 2000 について説明する。さらに、本研究で検討対象として用いる H.264/AVC 符号化技術について述べる。また、検討課題とする解像度変換の従来手法について説明する。

第3章 本研究で着目した時系列信号について述べる。まず時空間画像の概要について述べ、次にその特徴が用いられている関連研究について述べる。そして、時空間画像の符号化を提案し、これまでの研究において判明している問題点について述べる。

第4章 時空間画像の符号化により生じる影響について検討を行い、改善を試みる。まず、時空間画像の JPEG 2000 による符号化時に発生する問題について検討する。また、符号化効率の改善に寄与する要素について検討を行う。最後に、フレーム間予測を用いた時空間画像の符号化効率について検討する。

第5章 高速度カメラで撮影された高フレームレート映像に対して、時系列信号のフィルタリング処理により時間解像度変換を行う。まず、高フレームレート映像の特徴について説明する。次に、時空間画像のアプローチにより、高フレームレート映像に対する符号化について検討を行う。また、フィルタリングによる高フレームレート映像の時間解像度変換手法について検討を行う。

第6章 本研究の総括と今後の課題について述べる。

## 第2章

### 関連符号化技術

#### 2.1 はじめに

本章では，本研究で用いられる要素技術について述べる．まず，本研究の基盤となる JPEG 2000 符号化技術について説明する．次に，JPEG 2000 の拡張フォーマットである Motion-JPEG 2000 について説明する．さらに，本研究で検討対象として用いる H.264/AVC 符号化技術について述べる．また，検討課題とする解像度変換の従来手法について説明する．

#### 2.2 JPEG 2000

##### 2.2.1 概要

JPEG 2000 は，ISO/IEC JTC1/SC29/WG1 によって 2000 年 12 月に国際標準化された静止画像フォーマットである<sup>(7)</sup>．すでに広く普及していた JPEG よりも高い圧縮率と豊富な機能の実現と，様々な用途に対して一貫したアルゴリズムでの処理を目的とした．

##### JPEG 2000 のパート

JPEG 2000 には多くの要求仕様があり，パート 1 は JPEG 2000 の機能を最低限のレベルで実現するためのベースラインシステムとして位置づけられている．ビットストリームと復号器がパート 1 に準拠していれば，多様な画像端末の間でも互換性が確保されるようになっている．パート 2 は，パート 1 の拡張機能としてウェーブレット変換フィルタのユーザによる定義や，トレリス符号量子化を利用した高画質な量子化を実現できる一方，複雑度はパート 1 で規定されたスカラ量子化に比べて大きくなるため，あくまで JPEG 2000 の拡張オプションである．そのため，パート 1 の復号器はパート 2 の機能を満たす符号器が作成したビットストリームを復号できなくてよい．また，パート 3 は Motion-JPEG 2000 に関する規定であり，動画像を一連の静止画像のシーケンスとみなして，各画像フレームを JPEG 2000 で圧縮するための拡張機能である．Motion-JPEG 2000 については後で詳述する．

JPEG 2000 の各パートについて表 2-1 に示す<sup>(8)</sup>．

表 2-1: JPEG 2000 標準方式の各 Part

Part	Title	Detail
1	Core coding system	the syntax of the JPEG 2000 codestream and the necessary steps
2	Extensions	optional “value added” extensions
3	Motion-JPEG 2000	motion sequences of JPEG 2000 images
4	Conformance testing	information on compliance /conformance
5	Reference software	Java and C implementations are available
6	Compound image file format	JPM file format for document imaging
7	-	abandoned
8	JPSEC	security aspects
9	JPIP	interactivity tools, APIs and protocols
10	JP3D	volumetric imaging
11	JPWL	wireless applications
12	ISO Base Media File Format	common with MPEG-4
13	An Entry Level JPEG 2000 Encoder	an royalty and license fee free encoder with widespread applications

### JPEG 2000 の特徴

JPEG 2000 の基本機能としてコアシステムで実現された特徴を以下に示す<sup>(9)</sup>。

**高い圧縮効率** 全ての圧縮率において JPEG を上回る。特に高圧縮時に JPEG で顕著になるノイズ・歪みが目立たな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、全てのアプリケーションをカバーするのに十分な精度をもっている。ビデオで普及している全てのフォーマットに対応可能である。

**動画と音声の同期** 既存の Motion-JPEG では、動画と音声の同期法は規格化されていなかった。一方、DV ではフレーム単位に同期技術が確立されている。Motion-JPEG 2000 では、MJ2 file format によって動画と音声の同期を実現している<sup>(10)</sup>。

**可逆・非可逆圧縮** 可逆型 Wavelet フィルタ（非量子化）を使うことで、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を発揮する。

**プログレッション** JPEG2000 では5つの異なるプログレッションタイプを備えており、これらを使い分けることで非常に柔軟な画像のハンドリングが可能になる。

**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応用が考えられる。これは ROI (Region Of Interest) と呼ばれ、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、全部の画像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。

**エラー耐性** JPEG 2000, および Motion-JPEG 2000 では、その高い圧縮効率を利用することで、無線環境下での応用が期待されている。無線での画像伝送では基本的にエラー訂正コードが埋め込まれるものの、完全なエラー訂正には計算負荷・コストが高くなる欠点がある。そのため、符号器側で強力なエラー耐性能力をもっておく必要がある。また、インターネットの IP のプロトコルで画像のコードストリームを伝送する場合には、パケットロスに対する防備も十分に備えておく必要がある。

**ライセンス** 既存の JPEG と同様に、無償で使用できることが普及のために必要である。

### 2.2.2 符号化アルゴリズム

図 2-1 に JPEG2000 の符号化アルゴリズムの概略を示す<sup>(12, 13, 14)</sup>。以下、各要素技術について述べる。

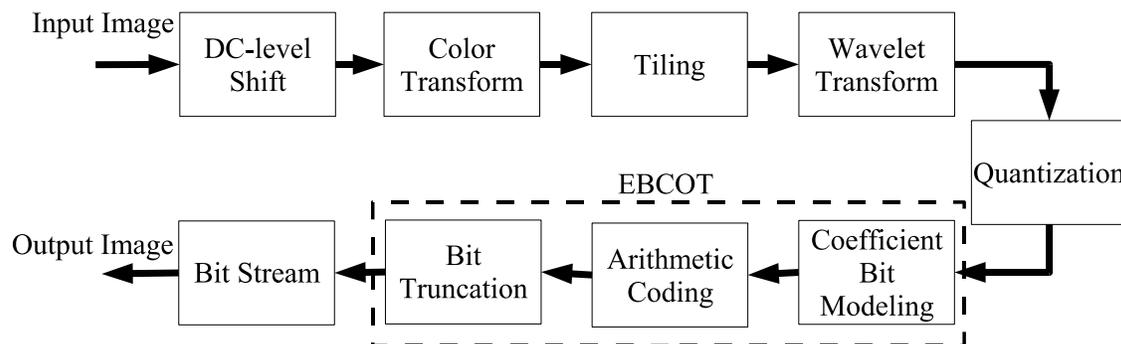


図 2-1: JPEG2000 符号化アルゴリズム

#### DC レベルシフト

入力信号が正の整数で与えられる場合に、ダイナミックレンジを半分にするすることで、ウェーブレット変換を効率的に行うことができる。逆に、復号時には各信号値にダイナミックレン

ジの半分を加算する処理を行う。入力信号が符号付き整数の場合には、このレベルシフトは行われない。

### 色変換

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(Irreversible multiple Component Transformation)の二つの変換が定義されている。それぞれの順変換と逆変換の定義式を以下に示す。

RCT

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} \lfloor \frac{R+2G+B}{4} \rfloor \\ R - G \\ B - G \end{bmatrix}$$

$$\begin{bmatrix} G \\ R \\ B \end{bmatrix} = \begin{bmatrix} Y - \lfloor \frac{R+2G+B}{4} \rfloor \\ Cb + G \\ Cr + G \end{bmatrix}$$

ICT

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.229 & 0.587 & 0.114 \\ -0.16875 & -0.33126 & 0.5 \\ 0.5 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0 & 1.402 \\ 1.0 & -0.34413 & -0.71414 \\ 0.5 & 1.772 & 0 \end{bmatrix} \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix}$$

この色変換処理はオプションであり、可逆変換を目的とするときはRCTを用いる、もしくは色変換を行わない。

### 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。しかし、高圧縮時にタイルの境界線が見えやすくなる問題がある。後述するMotion-JPEG 2000を想定すると、タイル間の境界に歪みが発生した場合に、静止画像では目立たなくても、連続フレームの動画像として観察すると、歪みが目立つ場合がある。

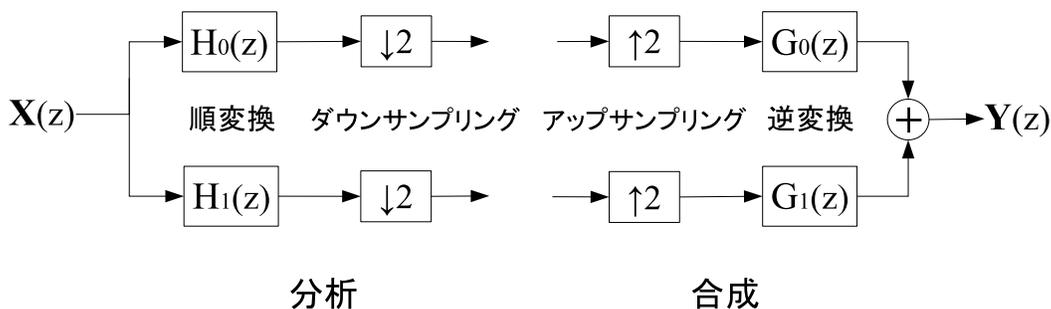


図 2-2: 2分割フィルタバンク



図 2-3: 2ステージ分解

ウェーブレット変換

JPEG2000 では、1次元の2分割フィルタバンクを繰り返し用いることで、2次元信号の離散ウェーブレット変換を実現している<sup>(11)</sup>。1次元の2分割フィルタバンクによる分析と合成の様子を図 2-2 に示す。

この分析処理を2次元の画像信号に水平・垂直方向に処理することで、LL, LH, HL, HHの4つの帯域に分割される。これを1ステージ分解という。他の帯域に比べて多くのエネルギーを持つ、水平・垂直方向ともに低域成分のLLを引き続いて分割することにより、オクターブ分割を行っている。図 2-3 に2ステージ分解した場合の帯域成分を示す。

ウェーブレット変換は、可逆変換を実現する整数型 DWT と非可逆変換を実現する実数型 DWT の2種類が存在する。JPEG2000 Part1 では、整数型フィルタとして変換係数が整数である5/3フィルタと、実数型フィルタとして変換係数が実数である実数型9/7フィルタの2種類が定義されている。整数型 DWT は、分割フィルタバンクにおけるローパスフィルタ (LPF:Low Pass Filter) のタップ数が5, ハイパスフィルタ (High Pass Filter) のタップ数が3の5/3フィルタが既定となっている。実数型 DWT は、LPF のタップ数が9, HPF のタップ数が7の9/7フィルタが既定とされている。整数型と実数型のそれぞれについて、分析側・合成側のフィルタ係数を表 2-2, 表 2-3 に示す。

表 2-2: 整数型 5/3 フィルタの係数

$i$	分析側		合成側	
	LPF $h_0(i)$	HPF $h_1(i)$	LPF $g_0(i)$	HPF $g_1(i)$
0	6/8	1	1	6/8
$\pm 1$	2/8	-1/2	1/2	-2/8
$\pm 2$	-1/8			-1/8

表 2-3: 実数型 9/7 フィルタの係数

(a) 分析側

$i$	LPF $h_0(i)$	HPF $h_1(i)$
0	0.6029490182363579	1.115087052456994
$\pm 1$	0.2668641184428723	-0.5912717631142470
$\pm 2$	-0.07822326652898785	-0.05754352622849957
$\pm 3$	-0.01686411844287495	0.09127176311424948
$\pm 4$	0.02674875741080976	

(b) 合成側

$i$	LPF $g_0(i)$	HPF $g_1(i)$
0	1.115087052456994	0.6029490182363579
$\pm 1$	0.5912717631142470	-0.2668641184428723
$\pm 2$	-0.05754352622849957	-0.07822326652898785
$\pm 3$	-0.09127176311424948	0.01686411844287495
$\pm 4$		0.02674875741080976

実数型は整数型に比べて処理演算量は多いが、符号化効率が高い。

また、表 2-2 のフィルタ係数を用いて、具体的に整数型 5/3 フィルタの分析側の伝達関数は

$$H_0(z) = (-z^2 + 2z^{-1} + 6 + 2z^1 - z^2)/8$$

$$H_1(z) = (-z^{-1} + 2 - z^1)/2$$

と表せる。また、合成側の伝達関数は、

$$G_0(z) = -H_1(-z)$$

$$G_1(z) = H_0(-z)$$

となる。

次に、実数型 9/7 フィルタの分析側の伝達関数を示す。表 2-3(a) の分析側フィルタ係数  $h_0(i), h_1(i)$  を用いて、

$$H_0(z) = \sum_{i=-4}^4 h_0(i)z^i$$

$$H_1(z) = \sum_{i=-3}^3 h_1(i)z^i$$

と表される。合成側の伝達関数は整数型 5/3 フィルタと同様である。

## 量子化

JPEG2000 Part1 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。ポスト量子化については後で述べる。スカラー量子化の量子化ステップサイズは、

$$\Delta_b = 2^{R_b - \varepsilon_b(1 + \frac{\mu_b}{2\Gamma})}$$

という式で計算される。 $\Delta_b$  はサブバンド  $b$  の量子化ステップ、 $R_b$  はサブバンド  $b$  のダイナミックレンジである。 $\varepsilon_b$  と  $\mu_b$  は逆量子化のためにデコーダに送られる、量子化ステップ表現のパラメータである。スカラー量子化には、各解像度レベルの全てのサブバンドごとに独立にステップサイズを計算する明示的な量子化 (expounded quantization) と、最低域のステップサイズから他のステップサイズを従属的に求める暗示的な量子化 (derived quantization) の 2 種類がある。前者は後者に比べ複雑であるが、より最適なステップサイズを選択できる可能性がある。

また、画像の特定の領域の画質を優先する ROI (Region Of Interest) をここで行うこともできる。

## EBCOT

ウェーブレット変換後の変換係数は、EBCOT(Embedded Block Coding with Optimized Truncation) と呼ばれるアルゴリズムによってエントロピー符号化される。これはウェーブレット変換により複数の帯域に分割された画像を、さらに

- コードブロック分割
- 係数ビットモデリング
- 算術符号化

という処理により、ビットレートに応じて画質を向上させる SNR スケーラビリティを容易に実現している。各処理について以下に述べる。

**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特異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 で規定された、学習型の2値算術符号化である。

**ビット切り捨て** 算術符号化されたビットストリームに対してポスト量子化を行い、画質への寄与が高い順に各コードブロックのデータを並べ替え、与えられたビットレートを超えるデータについては切り捨てられる。ポスト量子化のメリットは、符号量の制御を1パスで実現でき、レート制御のためのフィードバックが不要となることがあげられる。しかし、ポスト量子化では量子化の粗さを2のべき乗で制御することと等価であるため、低レート時に品質の微調整が困難となる。

なお、可逆符号化の際にはビット切り捨ては行われない。

## 符号ストリームの生成

算術符号化されたコードブロック単位のデータは、画像品質への寄与度に応じてエンベデッド符号を構成するため、レイヤと呼ばれる単位にグループ化される。各コードブロック

の対応するレイヤ情報は、パケットと呼ばれる単位にまとめられ、圧縮符号ストリームの基本的な単位となる。パケットの他に、データの区切りや符号化パラメータの伝達のために、マーカとマーカセグメントが符号ストリーム内に配置され、さらにメインヘッダが付加されて符号ストリームが形成される。

### 2.2.3 動向

同じ静止画圧縮技術である JPEG に比べ、符号化効率と画質の向上を主眼に JPEG 2000 は制定された。非可逆符号化時には JPEG と比較して、およそ 20 - 30 % の符号化効率の改善がみられた。また、低ビットレート符号化時に、JPEG では離散コサイン変換固有のブロック形状の符号化歪みが発生したが、JPEG 2000 では離散ウェーブレット変換が採用されたため、符号化歪みは“ボヤケ”として現れる。両者の視覚効果は異なるが、JPEG 2000 の優位性が際立つ形となっている。また、エンベデッド符号化により空間スケーラビリティを有するため、再符号化することなく、再生環境に応じて必要な符号量のみから復号することが可能となっている。

以上のような優位性から JPEG 2000 の普及が期待されたが、現時点では静止画の代表フォーマットとまで至っていない。これにはいくつかの要因が考えられる。第一に OS 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点が大いようである。JPEG という標準フォーマットが普及していることに加え、Microsoft が HD Photo という静止画ファイルフォーマットを開発し、次世代 OS である Windows Vista で標準搭載するという動きも影響しているようである<sup>(15)</sup>。HD Photo は、JPEG 2000 と比較して同程度の圧縮効率を実現し、演算量は JPEG 2000 よりも少ないとされる。高ダイナミックレンジまでサポートされ、可逆変換、非可逆変換が可能という、JPEG 2000 と同じ機能を有している。

広く普及しているデジタルカメラでの JPEG 2000 サポートも期待されるが、専用チップが必要な上、代表 OS である Windows に標準搭載されていない現状では、ファイルの扱いが困難である。そうした互換性の需要から、デジタルカメラでは未だに広く JPEG がメインフォーマットとなっている。

Windows 以外の代表 OS である Apple Mac OS X では、QuickTime の機能により JPEG 2000 が標準サポートされているが、Apple Safari 以外の Web ブラウザではサポートされていないため、Web ページでも使用できないのが現状である。Web ブラウザプラグインによる JPEG 2000 機能も提供されているが、処理が遅く実用的とは言えない。OS もしくは Web ブラウザ側での標準サポートが待たれる。

また、演算コストの増加分に見合うほどには符号量を抑えられないといった見方をする向きがある。さらには、近年の蓄積装置の大容量化から必ずしも高圧縮が求められてはいない。

以上のような観点から、未だ JPEG に取って代わるには至っていない JPEG 2000 であるが、符号化効率や機能に関する有効性は確かである。今後の動向にも注目されたい。

## 2.3 Motion-JPEG 2000

Motion JPEG2000 は、JPEG2000 の符号化方式を動画像に拡張したものである。Motion JPEG2000 は、MPEG のようにフレーム間の相関を利用する動き予測は行わず、フレーム単位で独立に符号化される。そのため、符号化効率は MPEG よりも劣るが、符号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。また、DCT 変換ではなくウェーブレット変換を用いているため、ブロック歪みも発生することなく空間的に滑らかな動画となる。また、静止画と同様に、整数型 DWT を用いることにより、可逆圧縮を実現することができる。一方で、ウェーブレット変換・逆変換において発生する量子化雑音の折り返し成分が原因のフリッカーと呼ばれるリングングノイズが現われるなどの問題もある<sup>(16)</sup>。

### 2.3.1 特徴

以下に Motion-JPEG 2000 の特徴について詳述する。

#### 動画静止画の共用化

静止画用と動画用とでは、目的やアプリケーションによって CODEC を使い分けているのが現状であり、静止画フォーマット、動画フォーマットともに乱立している。数多あるフォーマットの機能を満たし、動画と静止画に併用できる CODEC があれば、ユーザーにとってもメーカーにとってもメリットとなりうる。

#### 編集の容易性

Motion-JPEG 2000 では、連続する動画像の 1 枚 1 枚の画像を独立に JPEG 2000 で符号化するため、すべてイントラピクチャからなる。ピクチャ単位の復号ができるので、ピクチャの順番を入れ替えたり、切り出したり、あるいはスロー順方向再生、スロー逆方向再生、シャトルといった機能にも好都合である。MPEG-2 では、予測符号化が行われているため、圧縮効率は向上するものの、Motion JPEG 2000 に比べ編集が容易ではない。複数ピクチャ分の画像データを記憶するメモリも大きくなる。

#### 可逆・非可逆圧縮

JPEG 2000 Part-1 準拠である Motion-JPEEG 2000(Part-3) では、Part-1 と同じように可逆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を実現できる。また、同一の手法で非可逆圧縮を行うことも可能であり、可逆圧縮して得られたコードストリームを復号することなく、非可逆のコードストリームに直接変換することもできる。これは、MPEG などの従来の動画国際標準方式にはないメリットである。

## スケーラビリティ

JPEG 2000 の持つスケーラビリティ機能をそのまま使えるので、空間解像度レベルのプログレッション、画質レベルのプログレッションのほか、コンポーネントのプログレッションなど、豊富なプログレッションに対応できる。これらの機能は、WWW での高速かつ高効率なブラウジングや、ビデオ編集機でも有用ある。

## 解像度・ビットレート・フレームレートの自由度

JPEG 2000 では解像度の自由度が非常に大きく、規格上では<sup>231</sup>未満まで、水平・垂直サイズが対応可能になっている。また、ビット精度については、符号1ビット絶対値38ビットの計39ビット深度まで対応可能である。Motion-JPEG 2000 の場合には、本来が静止画の連続形式であることから、フレームレートは任意に設定可能である。このように、従来の動画符号化規格には存在しなかった自由度がある。

## アプリケーション

Motion JPEG2000 は、以下のようなアプリケーションへの適用が想定される。

- デジタルシネマ
- 放送用高画質デジタルシネマ記録機器
- デジタルカメラ
- 遠隔監視画像
- 高精細な医用画像，衛星画像

### 2.3.2 デジタルシネマ

JPEG 2000 のフォーマットが適用された動画アプリケーションとして代表的なものにデジタルシネマがあげられる。デジタルシネマとは、これまでフィルムを用いて制作・上映されていた映画の分野で、撮影段階からデジタルデータとして記録し、編集、配信、上映、保存までをフィルムレスでデジタルで完結させるシステムの総称である。これまでに蓄積された膨大な映画フィルムに対してもデジタルデータ化しアーカイブとして保存することで、ネットワークを通じて容易に配信を行うことが可能となり、映画産業全体に革新が起こることが期待される。

デジタルシネマは DCI (Digital Cinema Initiatives, LLC) によって標準化が行われ、2005 年 7 月最終使用が決定された<sup>(17)</sup>。この仕様は、将来にわたって利用可能な映像品質と拡張性を達成することを目指して策定された。デジタルシネマの仕様の概要について表 2-4 に示す。また、比較参考に国内の HDTV の仕様を併記する<sup>(18)</sup>。

表 2-4: デジタルシネマの規格概要

Title		Digital Cinema	HDTV
Video	Image Resolution	4096x2160, 2048x1080	1920x1080
	Frame Rate	4K: 24 [frame/s] 2K: 24/48 [frame/s]	60 [field/s]
	Compression Method	JPEG 2000	MPEG-2
	Bit Depth	12 [bit/component]	8 [bit/component]
	Color Space	XYZ	YPbPr
Audio	Sampling Rate	48 [kHz], 96 [kHz]	32 [kHz], 48 [kHz]
	Quantize	24 [bit]	16 [bit] ~
	Number of Ch.	max: 18 [ch]	max: 5.1 [ch]
Encryption		AES 128 [bit]	Multi2

映像データの圧縮には JPEG 2000 に基づくフレーム内符号化が採用されているが、動画に対する枠組みは Motion JPEG-2000 ではない。これには、JPEG 2000 Part-3 よりも相互運用性の高いファイルフォーマットとして MXF (Material Exchange Format) により、画像・音声データのパッケージングが施されることになる。

デジタル・シネマの標準仕様では、2 K モードと 4 K モードの 2 つが用意された。4 K のフォーマットは 35mm 銀塩フィルムに匹敵するほどの最高品質を提供する。また、2 K のフォーマットは、表示画素数こそ、HDTV とほぼ同じだが、階調数が強化され、グラデーション性能の優位を誇っている。

## 2.4 H.264/AVC

### 2.4.1 概要

H.264/AVC は、従来の MPEG 同様、動き補償予測付きフレーム間予測、変換・量子化、可変長符号化の組み合わせで構成され、高い符号化効率を提供する符号化方式である<sup>(19)</sup>。とりわけ、低レートや低解像度での符号化において、符号化効率の向上が大きく、MPEG-2/4 などでブロックノイズが多く発生するようなビットレートであっても、ブロックノイズの見えにくい画像に符号化することが可能である。各処理段階において、符号化モードの選択肢の拡大と徹底的な最適化による性能向上を追求し高い圧縮率が得られるが、従来方式に対して処理量が膨大となるため、圧縮性能を犠牲にしない処理負荷の低減技術の開発が課題とされている。

H.264/AVC は高い符号化効率を得られるため、高い圧縮率の必要なアプリケーションにおいて利用される。符号器には複雑かつ大量の演算処理が要求されるため、リアルタイムで動作させるには大規模の装置が必要となる。

### 2.4.2 符号化アルゴリズム

H.264/AVC では、MPEG-2/4 と比べ新しく特徴的な機能が盛り込まれた。H.264/AVC の符号化処理のブロック図を図 2-4 に示す。

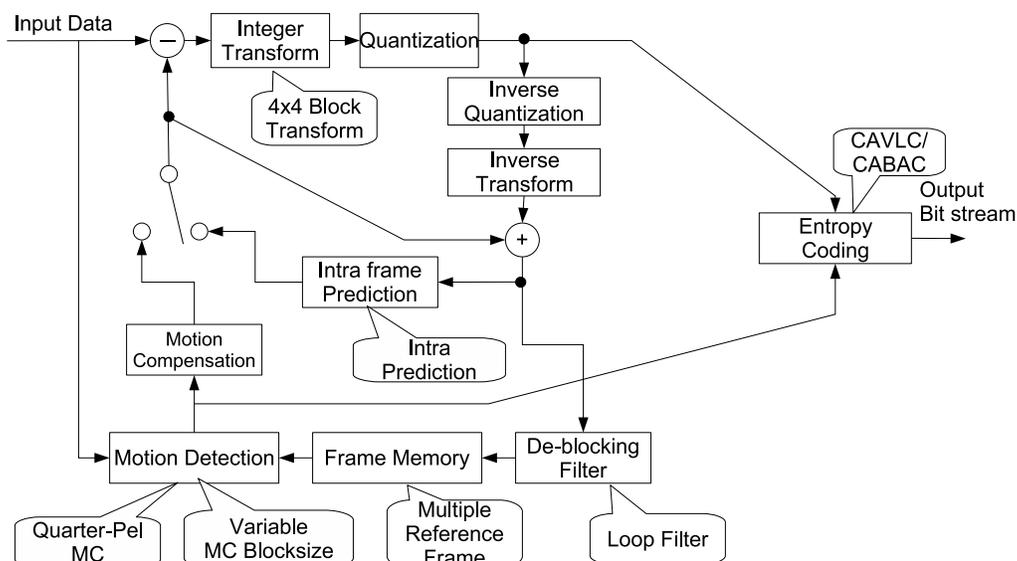


図 2-4: H.264/AVC 符号化アルゴリズム

個々の処理について、特徴的なものを以下詳細に示す。

#### 動き補償予測

通常、動画像信号ではフレーム間に高い相関があり、符号化済みのフレーム上の画素を用いて予測値との差分をとる手法がある。動き補償予測とは、被予測画素がすでに符号化されたフレームの中でどの位置に移動しているかをあらかじめ検出し、その動き量分補正した位置の画素を予測値として差分をとる方法である。フレームを画素ブロックに分割して、画素ブロック単位にブロックマッチングを行うことで動き量を検出し、動きベクトルとして、予測誤差の量子化結果とともに符号化される。動き補償のブロックサイズを小さくした方がより細かく予測を行うことが可能であるが、ブロックごとにベクトル情報を符号化しなければならないため、符号量が増えてしまう。そのため、どのブロックサイズを利用したほうが符号化効率が高くなるかを符号化時に判断する必要がある。動きベクトル探索は符号器によってのみ行われ、この動き補償予測処理は圧縮効率に最も大きな寄与を果たす要素である。

#### フレーム内予測

H.264 では、画素領域上でフレーム内予測処理を行い、符号量の削減を図っている。全てのプロファイルで 4x4 画素単位での予測、16x16 画素単位での予測が利用でき、High プロ

ファイルでは 8x8 画素単位での予測も追加された。フレーム内予測は上側と左側に位置する符号化済みのマクロブロックの画素を用いて予測を行う。16x16 画素単位での予測は 4 通り、4x4 画素、8x8 画素単位での予測にはそれぞれ 9 通りの予測モードが定義されており、発生符号量と符号化歪みを考慮してモードを選択する必要がある。

### フレーム間予測

MPEG-2/4 でも動き補償付きフレーム間予測は行われていたが、H.264 ではより多くのモードと複雑な演算処理の採用により予測効率を高めている。MPEG-2/4 では P ピクチャで 1 枚、B ピクチャで 2 枚に限られていた参照フレームであったが、H.264/AVC では、フレームメモリ内に蓄えられた複数のフレームを利用可能にし、予測効率を高めている。

また、動き補償を行うブロックサイズには 16x16、16x8、8x16、8x8、8x4、4x8、4x4 画素単位の 7 種類が設けられている。まず、16x16、16x8、8x16、8x8 の 4 種類のいずれかを選択し、8x8 画素単位を選択した場合には、さらに 8x8 のサブブロックごとに 8x8、8x4、4x8、4x4 画素単位のブロックサイズを選択するという階層構造がとられている。さらに、動き補償の精度が、1/2 画素精度から 1/4 画素精度まで向上したため、より高性能な予測が行えるようになった。

フレーム内予測、フレーム間予測いずれにおいても、予測モード数、参照フレーム数、可変ブロックサイズ、動き補償の画素精度の増加により、演算量が増加してしまう問題がある。動きベクトルを全探索すると符号化演算量が増加してしまう問題があり、演算量を削減する手法が多く検討されている<sup>(20)</sup>。

### 2.4.3 プロファイル規定

H.264/AVC では、主なアプリケーションを想定して、現在までに大きく 4 種類が定義され、動画の符号化時に使用可能な符号化ツールのセットが規定されている。

#### Baseline プロファイル

主に TV 電話などの双方向通信アプリケーションやモバイル AV アプリケーションなどの比較的 low-bitrate 領域での利用を想定したプロファイルであり、処理負荷が増大しないことを考慮して、最小限の符号化ツールに絞られたものとなっている。携帯端末向け放送でも当面本プロファイルが用いられることが決定されている。

#### Main プロファイル

標準 TV 解像度でのデジタル放送やケーブル TV、またはディスクレコーダなどのデジタル AV 家電での利用を想定したプロファイルであり、MPEG-2 での Main プロファイルに相当するものである。Baseline プロファイルに加え、両方向予測やインタレース画像対応

表 2-5: H.264 プロファイル構成

Coding Tools	Baseline	Main	High	Extended
I/P Slices	o	o	o	o
B Slices	x	o	o	o
Switching I/P Slices	x	x	x	o
CAVLC	o	o	o	o
CABAC	x	o	o	x
Interlaced Coding	x	o	o	o
Enhanced Error Resilience tools (FMO, ASO, RS)	o	x	x	o
Data Partitioning	x	x	x	o
8x8 vs.4x4 Transform Adaptivity	x	x	o	x
Quantization Scaling Matrices	x	x	o	x
Separate Cb/Cr QP control	x	x	o	x
Monochrome Video Format	x	x	o	x

の符号化ツールがサポートされるとともに、算術符号ベースの高性能な可変長符号化方式 CABAC(Context Adaptive Binary-Arithmetic Coding) の採用により、より高い圧縮性能を実現している。

### High プロファイル

HDTV などの高解像度画像やプロ用途のシステムへの適用をねらったプロファイルである。MPEG-2 で使用されていて、H.264/AVC では当初割愛されていた 8x8DCT と量子化マトリックスが本プロファイルでは使用可能となり、様々な映像入力に対してより柔軟な符号化モードの選択が可能になっている。本プロファイルは DVD フォーラムで検討されている HDTV 対応の次世代 DVD におけるビデオフォーマットの 1 つとして採用が決定している。

### Extended プロファイル

ストリーミングアプリケーションなどのオフラインエンコード処理を想定したプロファイルであり、基本的にすべての符号化ツールが使用可能となっている。しかし、デコーダの処理負荷低減のために CABAC のみ使用不可となっている。

以上、4 種類のプロファイルについて表 2-5 機能構成をまとめた。

#### 2.4.4 アプリケーション

H.264/AVC では、高い符号化効率を得られることから、高い圧縮率が必要とされるアプリケーションが当面の適用分野と目されている。一方、高い符号化効率を得るためには符号器に対して複雑かつ大量の演算処理が必要となるため、リアルタイムで動作させるにも、従来の MPEG 方式に比べると2～3倍の処理が必要となる。しかし、これは LSI、CPU の高集積化・高速化によりある程度実用可能なレベルまで達している。

日本国内の放送型サービスでは、地上デジタル放送の移動端末向けサービス、いわゆる“ワンセグ”にて H.264/AVC が採用された。また、蓄積型サービスとして、青色レーザを使用する次世代 DVD 規格 Blu-ray、HD DVD でも採用され、すでに製品化されている。双方向にリアルタイムサービスが必要となる対話型サービスでは、テレビ電話やテレビ会議用製品でも対応されており、PC 上で動くソフトウェアへの対応も進んでいる。

### 2.5 解像度変換

解像度変換には、サンプリングレートを上げることを目的としたアップサンプルと、下げることを目的としたダウンサンプルとがある。それぞれのレート変換には、数種類の方法があるが、本論文で検討を行うダウンサンプルについて述べる<sup>(21)</sup>。

#### 2.5.1 間引き

信号の解像度を落としたい場合に、所望のレートに基づいて単純に信号値の間引き手法が考えられる。動画像における時間解像度変換では、この手法は任意のフレームを落とすことに相当し、簡易な処理で実現できる。しかし、フレーム間引きを行うと、時系列上においてダウンサンプリングによるエイリアシングが発生し、視覚に影響を与えうるため、検討が必要である。

#### 2.5.2 デシメータ

ダウンサンプルで生じるエイリアシングを消去するために、周波数スペクトルにおいて、帯域制限を行い信号をカットオフする。デシメータに用いるデジタルフィルタを特にデシメーションフィルタと呼ぶ。デシメータを用いた帯域制限により、元の信号の情報を失っている点には注意が必要である。

#### 2.5.3 MCTF

MPEG によって、H.264 をベースにした次世代圧縮符号化方式 SVC の標準化が行われ、時間スケラビリティの確保に MCTF (Motion Compensated Temporal Filtering) が利用される。MCTF は動き補償時間方向フィルタと呼ばれ、時間方向サブバンド分割を再帰的に行うことで、低周波フレームと高周波フレームを生成する。低周波フレームを優先的に符

号化することで、効率よく符号化することが可能となる。

概要を図 2-5 に示す。

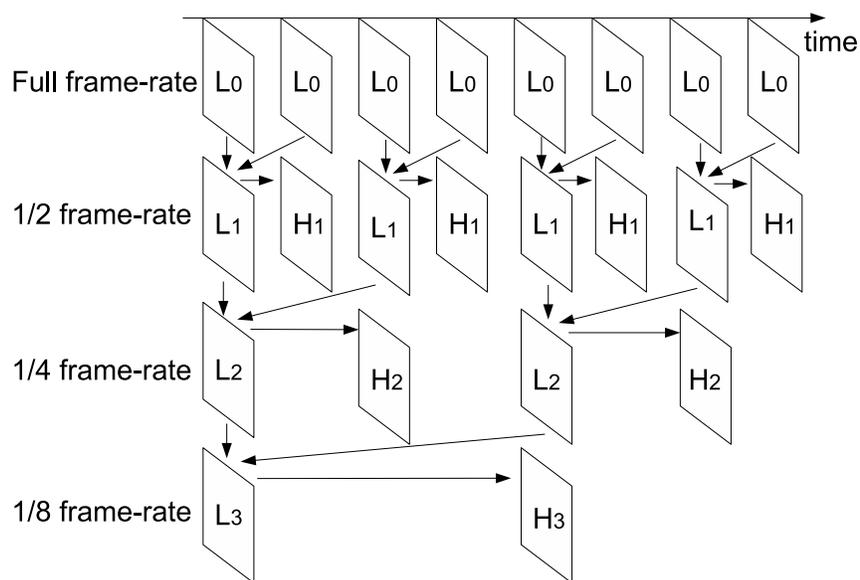


図 2-5: 動き補償時間方向フィルタ

サブバンド分割には 5/3 Wavelet Filter もしくは, Haar Wavelet Filter を選択できる。低周波フレームに対して再帰的にフィルタリングを行うことにより, 階層性を実現している。主にターゲットとされているのは, 30 [fps] からのダウンサンプリングである。

## 2.6 終わりに

本章では, 本研究の要素技術となる JPEG 2000 について説明した。また, 動画像符号化方式として本研究で検討を行う, Motion-JPEG 2000 及び H.264/AVC について述べた。最後に検討課題とする解像度変換の従来手法について説明した。

## 第3章

### 時系列信号

#### 3.1 はじめに

本章では、本研究で着目した時系列信号について述べる。まず時空間画像の概要について述べ、次にその特徴が用いられている関連研究について述べる。そして、時空間画像の符号化を提案し、これまでの研究において判明している問題点について述べる。

#### 3.2 時空間画像

##### 3.2.1 概要

時空間画像とは、2次元の空間座標からなる画像のいずれか1軸と時間軸とからなる画像である。そのため、選択されていない残りの空間軸上における時系列上での物体の動きを2次元画像として捉えることができる。図3-1に概要を示す。図3-1(a)は、時空間画像の定義を図示したものである。図3-1(b)に示された動画像が与えられた場合、ある空間軸における、時系列上の物体の動きがそれぞれ図3-1(c)のX-t平面、図3-1(d)のY-t平面のように得られる。なお、空間2軸と時間軸からなる3次元のデータ構造を時空間ブロックと定義する。時間方向のサンプル数は任意である。

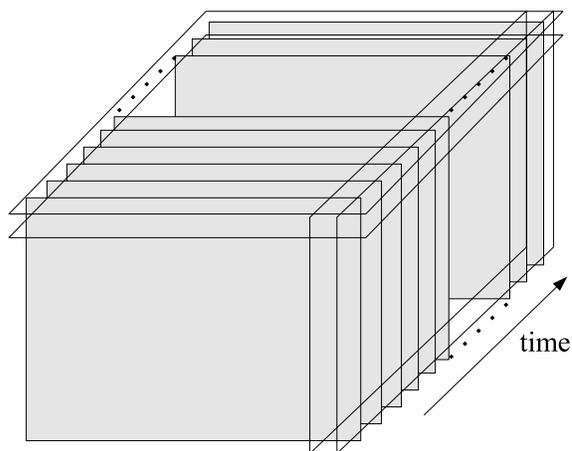
任意の空間軸における物体の時間的な動きを、画像として扱うことができる。また、空間情報と一律に捉えることが可能なため、3次元の信号の同列での比較が可能となる。

一般に、カメラで撮影されたシーケンスは、隣接するフレーム同士は類似しており、時間的な相関が高いという特徴があり、フレーム間予測処理はこのことに基づいて行われているが、画像という概念ではない。

##### 3.2.2 関連研究

映像処理のあらゆる分野において上記の時空間画像の特徴を有効に生かした研究が行われている。

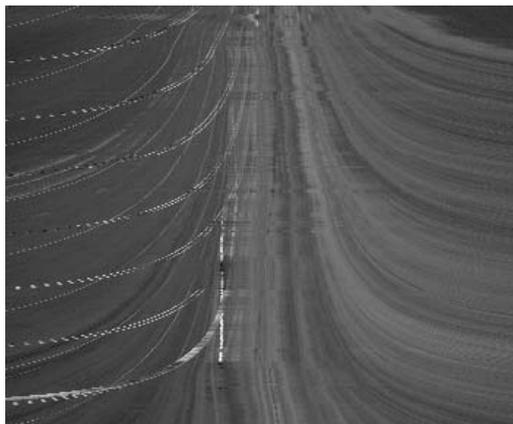
オブジェクトベースの符号化を目的に watershed 法を時空間に適用することにより、動画像から動オブジェクトを抽出する研究がなされている<sup>(23)</sup>。MPEG-4<sup>(24)</sup>で採用されているオブジェクト符号化には、動画像からオブジェクトを抽出することが必要不可欠であるが、



(a) 時空間画像の定義



(b) 空間画像 (XY)



(c) 時空間画像 (Xt)



(d) 時空間画像 (Yt)

図 3-1: 時空間画像

各フレームで輪郭形状が大きく異なると、符号化効率が非常に悪くなるため、輪郭は時間方向に出来るだけ滑らかになることが望ましい。そこで、境田らの研究では、watershed と呼ばれる領域成長法を時空間に拡張し、輪郭付近に適用することで動オブジェクトの抽出に有効となりうる手法を提案している。また、同様のアプローチで映像修復を目的として、フレーム同士の時間的な相関を利用することにより映像中のオブジェクトの消去もしくは補完を行う研究がなされている<sup>(25, 26, 27)</sup>。時間相関から映像中のオブジェクトを複数のレイヤに展開し、特定レイヤを除去することでオブジェクトを消去する手法が提案されているが、背景ないしはオブジェクトが移動しない映像に対してはこの手法は不十分である。そこで庄司らは、時間特徴と空間特徴を用いた背景補完を交互に組み合わせることにより、背景が移動しない場合であってもオブジェクトの消去を実現している。1フレームのみに空間特徴を用いた背景補完を適用し、他のフレームに対しては時間特徴を用いた補完を行うことによ

り、処理時間の軽減とともに、時間軸上での急激な色値の変化を防ぐことも可能となり、視覚的に自然な映像が得られる。

物体検出の分野として、各フレームから物体に関する有為な空間上の1次元データを抽出し、時系列に並べて得られる時空間画像を処理することによる物体の検出法が提案されている<sup>(28)(29)</sup>。谷口らは、検出すべき特徴データを投影することで、処理として物体の形状に対する依存度を弱め、各フレームでの物体抽出率を向上している。また、物体の移動に関する時間情報を領域として捉えることで、抽出率を高め、結果として環境変動に対してもロバストになる。本多らは、車載カメラによる映像をカーナビゲーションシステムに適用し、実写ベースでの案内表示を可能にしている。横断歩道などの道路上の白線を精度高く抽出するために、カメラからの映像を時空間画像として処理し、時間的に連続したその画像から目印となる白線を検出し位置補正を実現している。

阿久津らによって映像解析のために投影法を用いて時空間投影画像として可視化し、フィルタ処理により時空間画像内の動き情報を“流れ”として捉えるといった研究がなされている<sup>(30)</sup>。得られた情報からカメラワークを中心投影でモデル化することで、パンニング、ズームングといったカメラワークを抽出し、よりインタラクティブな映像ハンドリングを可能にする手法を提案している。また、岡本らは、固定長の時空間画像から、複数のショットを含むパッケージであるか単一のショットを含むパッケージであるか判定し、同時生起行列を用いて時間軸方向の特徴を陽に表現することで、パッケージを単位としてシーンごとにクラスタリングする手法を提案している<sup>(31)</sup>。

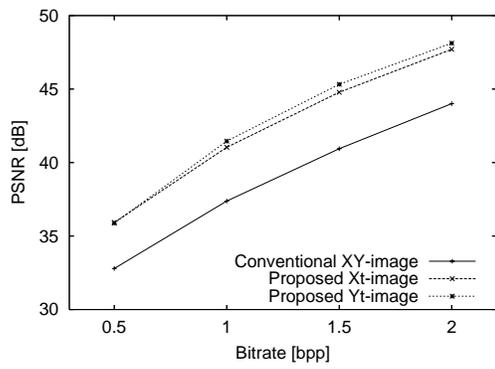
### 3.3 時空間画像の符号化

#### 3.3.1 提案手法

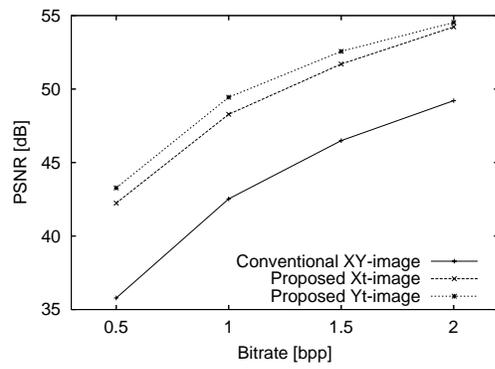
これまで、以上のような性質を持つ時空間画像を1フレームとして符号化し、時空間ブロック全体での符号量を抑えることを提案してきた<sup>(32)</sup>。JPEG 2000を用いて時空間画像として符号化したシーケンスを復号し、各ビットレート毎の平均PSNRを図3-2に、時間軸に沿ったフレーム毎のPSNRを図3-3に示す。図3-2は、4つのシーケンスに対して、XY、Xt、Ytの3種類の平面に対して非可逆変換を行っている。図3-3は、Xt、Yt平面に対してそれぞれ可逆・非可逆符号化した結果である。以降、JPEG 2000符号化には、参照ソフトウェアKakadu Ver. 5.2を用いて実験を行う<sup>(33)</sup>。

時空間ブロックを同じ符号量に揃えた場合でも、符号化効率が改善されていることが分かる。時系列上の相関を画像の1軸として利用した符号化が有効であることが証明される。この手法による処理時間は、時間に沿った符号化と比較して、時空間画像に展開するのに要する時間が増える程度であり、その時間はわずかである。また、復号する際にも再配列に要する時間が加えて必要となる。

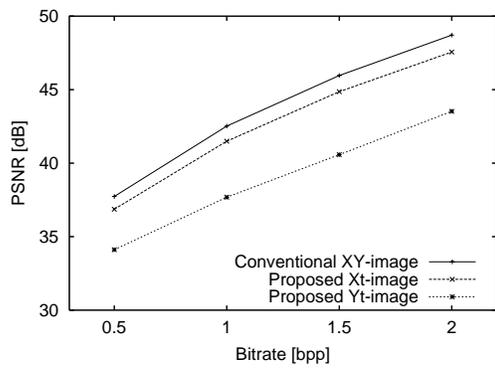
ストリーム再生される場合には、利用した時系列の長さ分だけのバッファ時間及びメモリ量を必要とするが、単位とする時間の長さを変えることでそれらを制御することが可能である。しかし、符号化に利用可能となる時間相関は時空間ブロックの時間の長さに比例すると



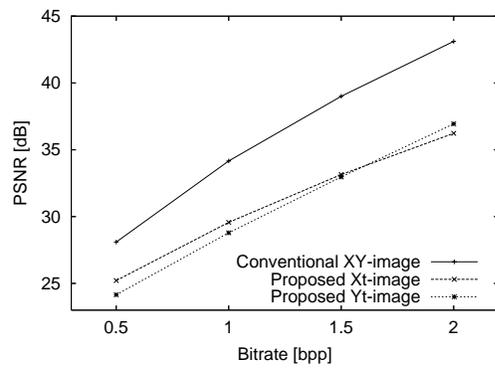
(a) table



(b) news



(c) canoe



(d) stefan

図 3-2: 時空間画像符号化による平均 PSNR

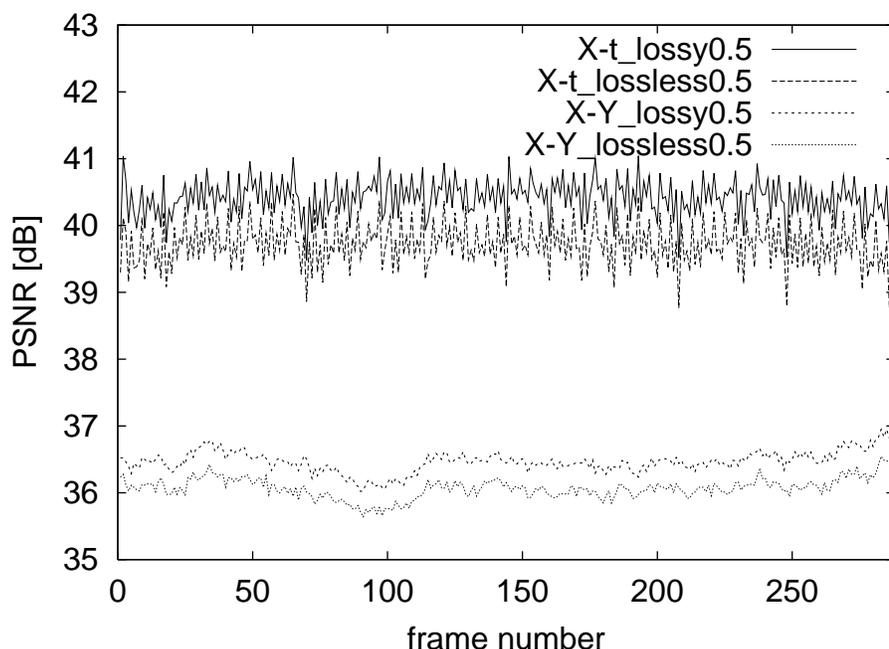


図 3-3: Hall Monitor (bit rate: 0.5[bpp])

考えられ、適切な時間長で区切る必要がある。

### 3.3.2 問題点

#### 符号化効率の揺らぎ

時間の相関を利用した符号化により、シーケンス内での画像品質の変動を抑制し、等価的に VBR 符号化のようにフレーム間の品質変動を抑制する効果が期待される。しかし、図 3-3 から、フレーム毎に細かく PSNR が変動していることが確認できる。全体的に高い PSNR であるか、変動幅が小さければ主観的に問題はないが、PSNR が低い映像でのフレーム毎の揺らぎは、主観品質を損ない、VBR 符号化の役割を果たさなくなる可能性がある。なお、時間に沿った通常の符号化でも時系列上で PSNR の変動が生じているが、これはシーケンスの複雑度によるものである。

#### 符号化効率の急激な低下

また、図 3-4 は特定の 1 フレームを時間に沿って並べた、全く動きのないシーケンスに対して、可逆及び非可逆 Wavelet 変換を行い、ビット切捨て及びスカラー量子化を行った結果である。ビット切捨てを行った場合に、あるフレーム毎の SNR の急激な低下が確認される。

時系列上で変動しないはずの PSNR がある規則をもって低下しているであろうことが確認される。上述した揺らぎの問題と同じく、復号・再構成後の視覚的な劣化につながる恐れ

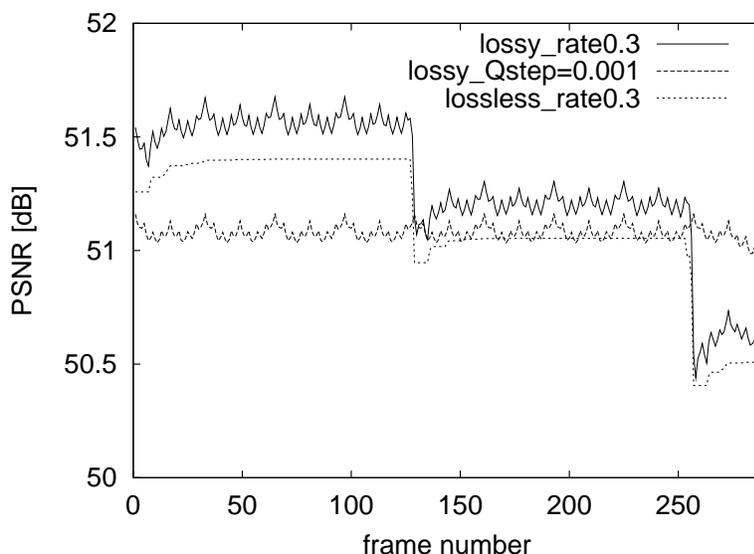


図 3-4: 時間的変動なし時空間画像 PSNR

があり，原因を解明する必要がある．

#### 符号化効率の悪化

シーケンスによって，提案手法の符号化効率の傾向が変わり，必ずしも提案手法が有効とは言えない．どのような要素が符号化効率に影響を与えているか解析する必要がある．

シーケンス内のカメラの動き，もしくはオブジェクトの動きが小さいことは，時間方向の相関が高いのと等価であり，時空間画像による符号化が有効に機能することは容易に期待されるが，定性的な解析を行う必要がある．

また，時空間画像という観点からインター符号化を行った例がなく，これまでの時空間画像のイントラ符号化と同様に，インター符号化でも符号化効率の改善が見込まれるか検討を行う必要がある．

### 3.4 高フレームレート映像

#### 3.4.1 概要

高速度カメラで撮影されたシーケンスについて述べる．一般に動画は，30 [frame/s]，60 [field/s] のものが広く普及しているが，デバイス技術の向上に伴って，より時間間隔を短く撮像するカメラも存在する．100 [frame/s] 程度のもことから，2億 [frame/s] を誇るものまで存在する<sup>(34)</sup>．高フレームレートの値に対する定義は曖昧であり，5 [frame/s] 程の低フレームレートに対比させた値として，30 [frame/s] 程のものを言うこともある．しかし，本研究では，一般に広く普及している表示系で視聴されうるフレームレート，60 [frame/s] より高

い値を高フレームレートと定義し、そのようなフレームレートで撮影された映像を高フレームレート映像と定義する。

撮影の時間方向のサンプリング間隔が密になるため、物体の時系列上における動きがより細かく記録される。こうして撮影された映像は、スロー再生によって物体の動きを追従するだけでなく、オブジェクト解析、信号解析などの研究に利用される。

### 3.4.2 問題点

フレームレートが向上すると、フレーム間の相関も高まることが期待される。しかし、一定時間間隔中におけるフレーム数が増え、それに伴い信号量も増加する。そのため、フレーム間相関を効率的に利用した符号化が求められる。

### 3.4.3 時間解像度変換

高フレームレートで撮影されたシーケンスを、任意の表示システムで見ると、それぞれのスペックに合わせてフレームレートを変更する必要がある。また、人間の視覚特性により、感知できるフレームレートには限界があり、それ以上は冗長である。また、それだけのレートを再現可能な表示デバイスも存在しない。

そこで、高フレームレートから 30 フレームなど一般のデバイスで表示可能なレートまで下げることが行われているが、単純に間引いただけでは、動きの滑らかさが損なわれ、またサンプリング定理が満たされておらず、エイリアシングの発生が懸念される<sup>(35)</sup>。

また、2章で MCTF について述べたが、SVC で採用された MCTF がターゲットとするのは 30[fps] からのダウンサンプリングであり、高フレームレート映像に対する有効性については検討の余地がある。

動画像の 3 次元データにおける時系列信号を画像の 1 軸として時空間画像を扱った場合、その解像度変換は時間解像度変換と等価である。時空間画像からのアプローチによる、時間解像度変換手法について第 5 章で検討する。

## 3.5 おわりに

本章では、本研究で着目した動画像における時系列信号について述べた。まず時空間画像の概要について述べ、次にその特徴が用いられている関連研究について説明した。次に時空間画像の符号化のアプローチについて説明した。最後に時系列信号の観点から、高フレームレート映像について述べた。

## 第4章

### 時空間画像符号化の改善手法

#### 4.1 はじめに

本章では、前章で述べた時空間画像の符号化により生じる諸問題について考察を行い、符号化効率の改善のための提案手法について説明する。最後に実験によりこの手法の有効性を示す。

#### 4.2 実験と考察

##### 4.2.1 PSNRの揺らぎについて

図3-3の結果より、時空間画像を符号化し、ビット切り捨てによるレート制御を行った場合、可逆、非可逆ともにPSNRが周期的に揺らいでしまっていることがわかる。これは符号化によって損失された部分が軸を変えたことにより現れたものである。このPSNRの周期的な揺らぎがWavelet変換特有の問題ではないかと考察し、それを実験で確認する。図4-1は垂直方向は一定で、水平方向のみに変化のある画像である。カメラの動きがなく、オブジェクトの動きもないシーンの時空間画像は、このように時系列上では変化のない縞状の画像となる。図4-1をJPEG2000で符号化した際のウェーブレット係数は、図4-2のように現れる。垂直方向には変化のない画像であるため、変換係数は図4-2に見られるように水平成分の低域に集中し、いずれの分解レベルにもLH及びHH成分の係数は存在しない。



図 4-1: 時間的変動なし時空間画像



図 4-2: 時間的変動なし時空間画像 Wavelet 係数

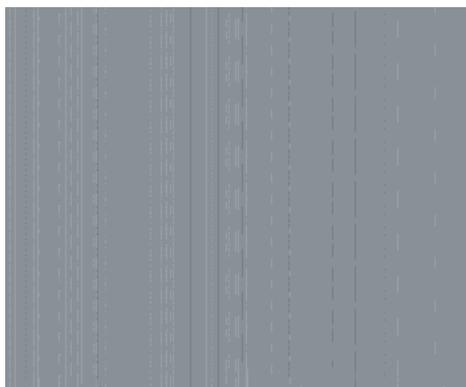


図 4-3: 非可逆変換の差分画像 (時間的変動なし時空間画像)

実数型の  $9/7$  フィルタを用いたウェーブレット変換は非可逆変換となるため、原信号には戻せない一種のノイズを伴ってサブバンド分割される。さらに、LL 成分はノイズと共に再帰的に分析され、サブバンド信号が得られる。そのようにして求められた、ノイズを含む信号から元の解像度に合成した画像を、図 4-3 に示す。水平方向では歪みが発生する場所は不規則であるが、垂直方向にはいくつかの周期を持って符号化ノイズが発生していることが確認される。

整数型の  $5/3$  フィルタを用いたウェーブレット変換は可逆変換となるため、ウェーブレット係数自体は符号化雑音を持つことはない。しかし、各サブバンドレベルにおける H 成分が量子化されることによって、合成される際に損失が生じる。

LL 成分は再帰的に分解されるため、高域成分が損失した際に、Wavelet の基底に倣った欠落が生じる。符号化ノイズが発生する場合は  $2^p$  ( $p = 1 \sim N_L$ ; 整数,  $N_L =$  分解数) を周期に持つ。こうした周期的な損失が各フレームで発生するために、時系列上での SNR の揺らぎになると考えられる。水平・垂直方向に動きがある場合にも同様の現象が発生していると考えられ、動きが少ないシーンでより顕著に確認される。

上述した現象はサブバンド分割を行う JPEG2000 の構造的な問題であり、このような符号化損失は時空間画像に限らず一般的な画像でも発生しうるが、空間内においてはわずかであり、視覚的には時空間画像の符号化においては、残りの空間 1 軸分だけ蓄積し、結果的に時系列上における SNR の揺らぎにつながる。

こうした現象はサブバンド分割を行う JPEG2000 の構造的な問題であり、JPEG2000 を用いた場合には不可避である。そこで符号化する時空間画像を図 4-4 のように取得する手法が考えられる。

時空間画像の時間軸における画像取得開始点を 1 フレーム分ずつ交互にずらすことにより、2 の周期を持つ符号化雑音による PSNR の揺らぎが打ち消されることが予想される。前後の時空間ブロックの境界でのサンプリングポイントをオーバーラップさせることにより、選択されないポイントをなくすことが可能である。また、従来手法では時空間ブロック毎に完全に閉じられていたため、時空間ブロック毎に品質の変動が生じる恐れがあったが、オー

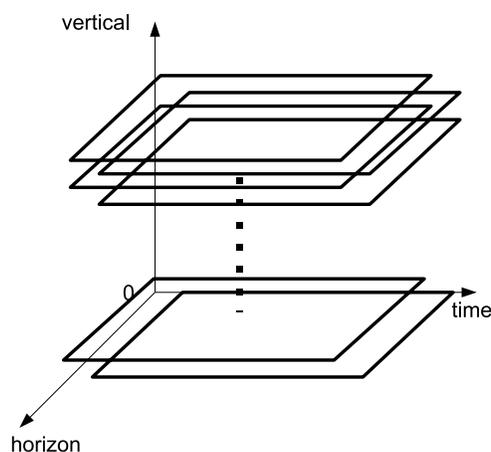


図 4-4: 提案手法 (ジグザグスキャン)

オーバーラップさせることにより、ブロック間でより滑らかに移行することが期待される。結果として品質変動を抑えた符号化を実現している。なお、この処理によるメモリ使用量は、従来手法に比べオーバーラップする部分が増える程度である。

以上の方法で得た時空間画像による符号化効率を図 4-5 に示す。

SNR の揺らぎの高低が平均化され、周期的な変動が抑制されていることが確認される。また図 4-6 に示した Table の 130 フレーム辺りはシーンの変化点であり、以降のシーンはやや複雑となる。シーケンス内で画像の複雑さが変動する場合、従来の時間に沿ったイントラ符号化では、PSNR が極端に落ちてしまっており、その後向上することは見られない。しかし、提案方式では、シーンの境界では PSNR が大きく低下してしまうものの、その後、ほぼ元の水準まで PSNR が戻っていることが確認される。時空間画像の符号化が有効で時空間ブロックでの平均 PSNR が向上するようなシーケンスにおいては、シーンの変化後も高い PSNR が得られることがわかる。数学的には、開始点が偶数の時空間画像と奇数の時空間画像を時空間ブロック内で同じ枚数だけ符号化すれば、SNR 上の揺らぎは解消されることが期待できるが、交互に配置することにより、劣化が空間内で偏った場所に現れるのを防ぐことが可能と考えられる。上記の実験の時空間ブロック全体での PSNR の平均値及び分散値を表 4-1 にまとめる。

提案手法により、奇数分だけシフトさせた場合、PSNR の標準偏差が小さくなり揺らぎが抑えられていることが確認できる。なお、提案手法により揺らぎを抑えると、わずかではあるが平均 PSNR が下がってしまうことが確認される。また、Hall Monitor のようにカメラが固定されており背景がほとんど変化しないようなシーケンスの場合、各フレームの複雑度はほぼ一定となり、Motion-JPEG 2000 による符号化でもフレームごとの変動が小さいことが確認される。

以上より、参照画素を 1 ずらすことにより、周期 2 の揺らぎが解消されることを確認した。しかし、こうした周期的な変動は周期 2 のみでなく、2 の階乗ずつ発生していることを述べた。ここでは、1 フレームおきに発生し、もっとも視覚的影響の高いであろう、周期 2 の揺

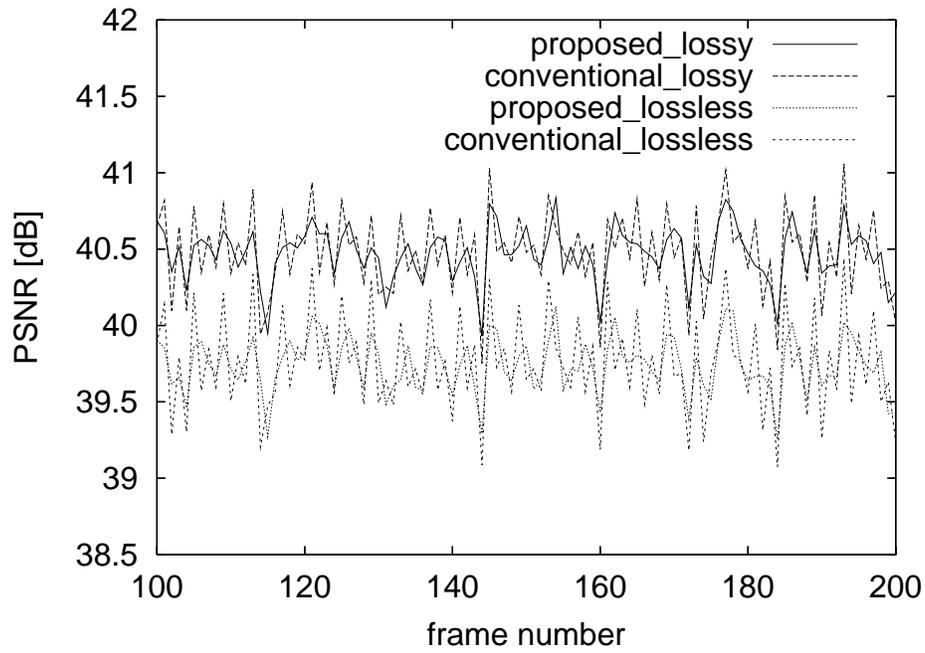


図 4-5: Hall Monitor PSNR

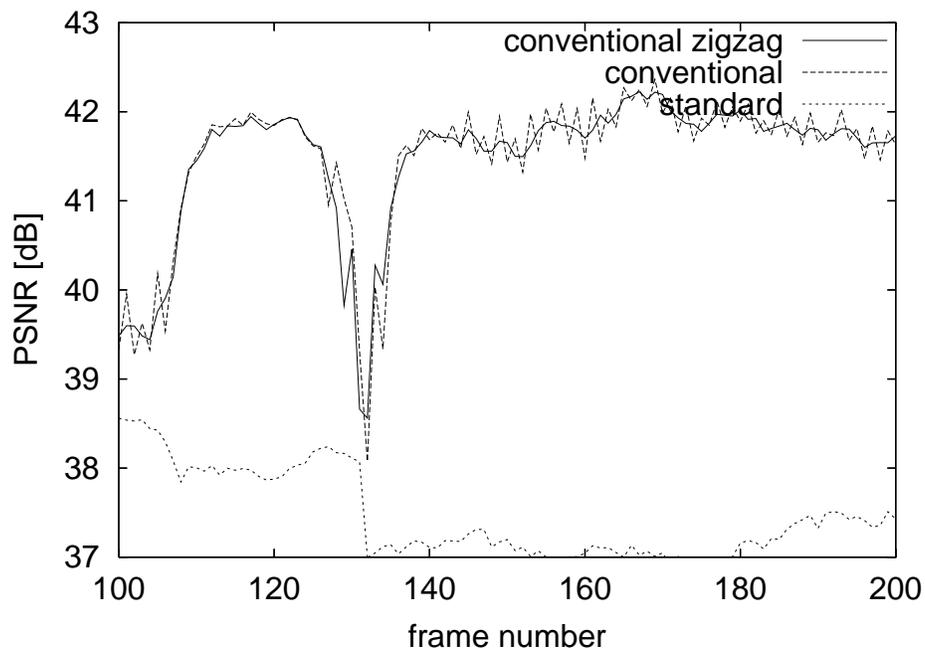


図 4-6: Table PSNR

表 4-1: PSNR の平均 / 標準偏差 (Hall Monitor 1.0[bpp])

		Lossy		Lossless	
		Mean	Variance	Mean	Variance
Motion-JPEG 2000		41.926	0.0135	41.001	0.0102
Conventional		43.745	0.1359	42.665	0.2607
Proposed	shift=1	43.734	0.0124	42.637	0.0356
	shift=2	43.741	0.1097	42.651	0.1789
	shift=3	43.724	0.0126	42.630	0.0305
	shift=4	43.742	0.1211	42.661	0.2394
	shift=5	43.730	0.0118	42.630	0.0293

らぎについて検討を行った。

また、シーケンスを2つの時空間ブロックに分割した際の符号化効率を図4-7に示す。

周期的な揺らぎが平滑化された結果、時空間ブロックの境界でのPSNRの変動が抑えられていることが確認される。

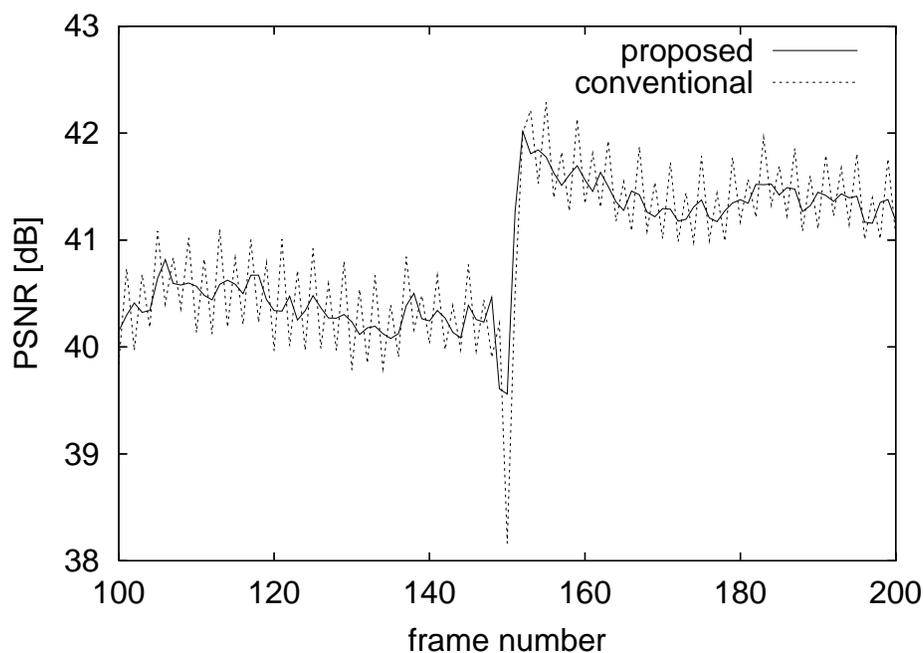


図 4-7: canoe PSNR

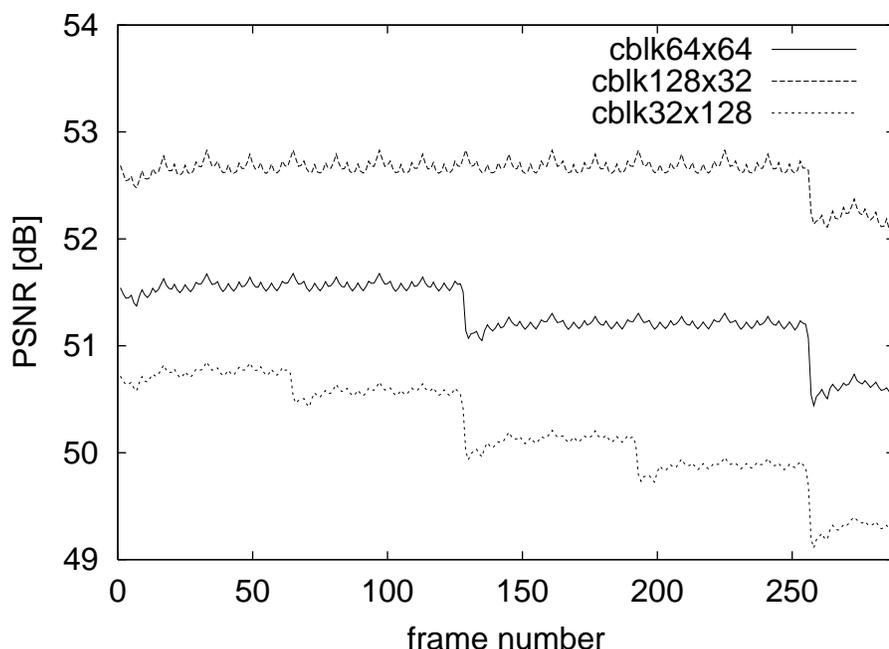


図 4-8: コードブロックサイズ

#### 4.2.2 SNR の急激な低下について

JPEG 2000 は一般には EBCOT によって  $64 \times 64$  のコードブロックに分割され、それ以降の符号化処理をコードブロック毎に独立に行われるよう規定されており、高域成分のコードブロックが切捨てられた後、より低域にまで切捨てが及んだことにより、コードブロックサイズで歪みが生じていると考えられる。

周期的な揺らぎの発生と同様に、1枚の時空間画像においてはわずかな歪みであっても、残りの空間軸分だけ蓄積されることにより、時系列上において、急激に低下する場合がある。

コードブロックサイズを変化させて符号化した場合のフレーム毎の PSNR を図 4-8 に示す。コードブロックサイズの 2 倍のフレーム間隔毎に SNR が低下していることが確認される。JPEG 2000 では、サブバンド分割されたウェーブレット係数を、EBCOT によってコードブロックに分割し、それ以降のビット切捨てを含む符号化処理はコードブロック毎に独立に行われる。そのため、コードブロック毎の切捨てによって、コードブロックサイズの倍数毎に境界となって現れると考えられる。

こうした現象も周期的な揺らぎ同様、JPEG 2000 を用いる際の構造的な問題である。

そこで、コードブロックサイズを変化させた際の符号化効率を図 4-8, 4-9 に示す。なお、実験に用いるシーケンスはシーケンス内の動きの影響を取り除くため、任意の 1 フレームを 288 枚並べたものを用いた。可逆、非可逆いずれにおいても、ある一定の周期ごとに PSNR が急激に低下していることが確認される。このことは、符号化する際に時間方向の信号がコードブロックサイズごとに分割されることに起因する。コードブロックはサブバンド分割された信号に対して作成されるため、復号される際に 1 つのコードブロックが影響を与える

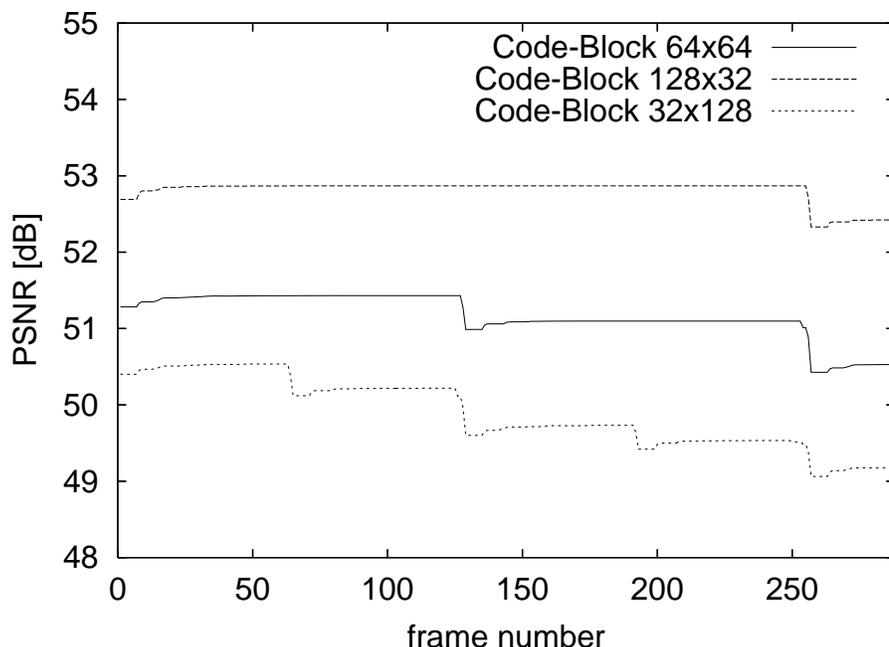


図 4-9: コードブロックサイズ (Lossless)

範囲はコードブロックサイズの水平幅，垂直幅のそれぞれ2倍となる。

また，コードブロックの境界付近においては，境界でない部分よりも，やや低下していることも確認される．コードブロックの境界付近にノイズが集中することが考えられる．

#### 4.2.3 シーケンスへの依存度について

これまで，時空間画像の符号化が有効であることを述べたが，図3-2に示された通り，シーケンスによっては低下する場合もある．どのような要素が符号化効率に影響を与えるかを考察する．

時間変動が少ない場合，時空間画像の方が空間に閉じられた画像より符号化しやすいことが想像される．これを定性的に求める．時空間ブロック全体での自己相関関数を求める．1次元信号  $s(t)$  ( $0 \leq t \leq N-1$ ) の自己相関関数  $r(l)$  は

$$r(l) = \frac{1}{N-l} \sum_{t=0}^{N-l-1} [s(t+l) - \tilde{s}][s(t) - \tilde{s}]$$

$$\tilde{s} = \frac{1}{N} \sum_{t=0}^{N-1} s(t)$$

という式で表される．これを3次元信号  $z(x, y, t)$  ( $0 \leq x \leq X, 0 \leq y \leq Y, 0 \leq t \leq N$ ) まで拡張した自己相関関数  $\Phi(p, q, r)$  の式を以下に示す．

$$\Phi(p, q, r) = \frac{1}{(X-p)(Y-q)(N-r)} \sum_{t=0}^{N-r-1} \sum_{y=0}^{Y-q-1} \sum_{x=0}^{X-p-1} [z(x+p, y+q, t+r) - \tilde{z}][z(x, y, t) - \tilde{z}]$$

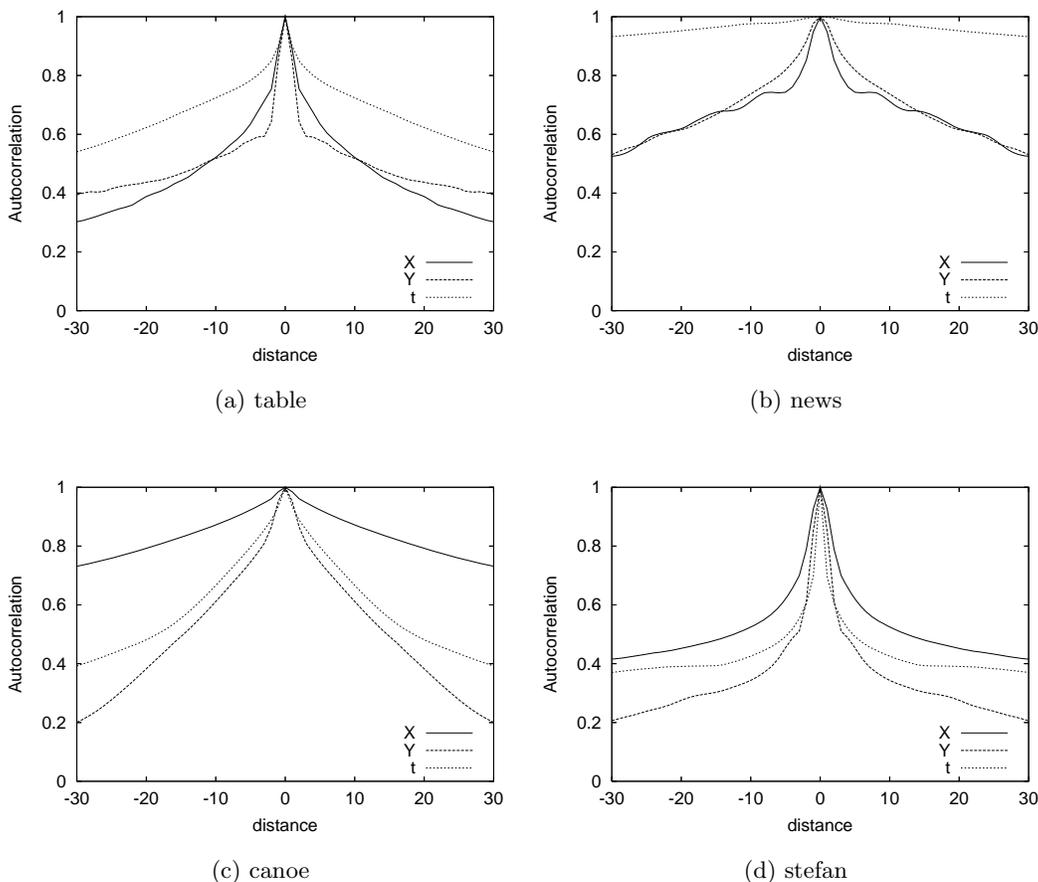


図 4-10: 1次元の自己相関関数

$$\tilde{z} = \frac{1}{X \times Y \times N} \sum_{t=0}^{N-1} \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} z(x, y, t)$$

上記の式より，各  $p, q, r$  を変化させ，それぞれ 3次元の自己相関関数を求めた際の時空間ブロックにおける平均自己相関関数を図 4-10 に示す．

図 3-2 及び図 4-10 より，時空間画像の符号化の有効性は相関の高さに依存していると考えられる．時空間画像の符号化の有効性が大きかった (a)table, (b)news はやはり時間軸の自己相関関数が高く，逆に (c)canoe, (d)stefan は  $t$  軸の自己相関関数が (a), (b) に比べ低いため，時空間画像の符号化に向いてないことが示される．符号化を行った 2次元の相関についても図 4-11 に示す．

1次元の相関関数のそれぞれの軸の値の中間の値をとり，この関数の特に距離が 0 近辺の，遠くない距離での相関の値が，時空間画像符号化の有効性について示していると考えられる．

以上の結果より，時空間画像の符号化の有効性はシーケンスへの依存度が高いため，それぞれの相関の強さを考慮し，適応的に選択する方法が有効であると考えられる．

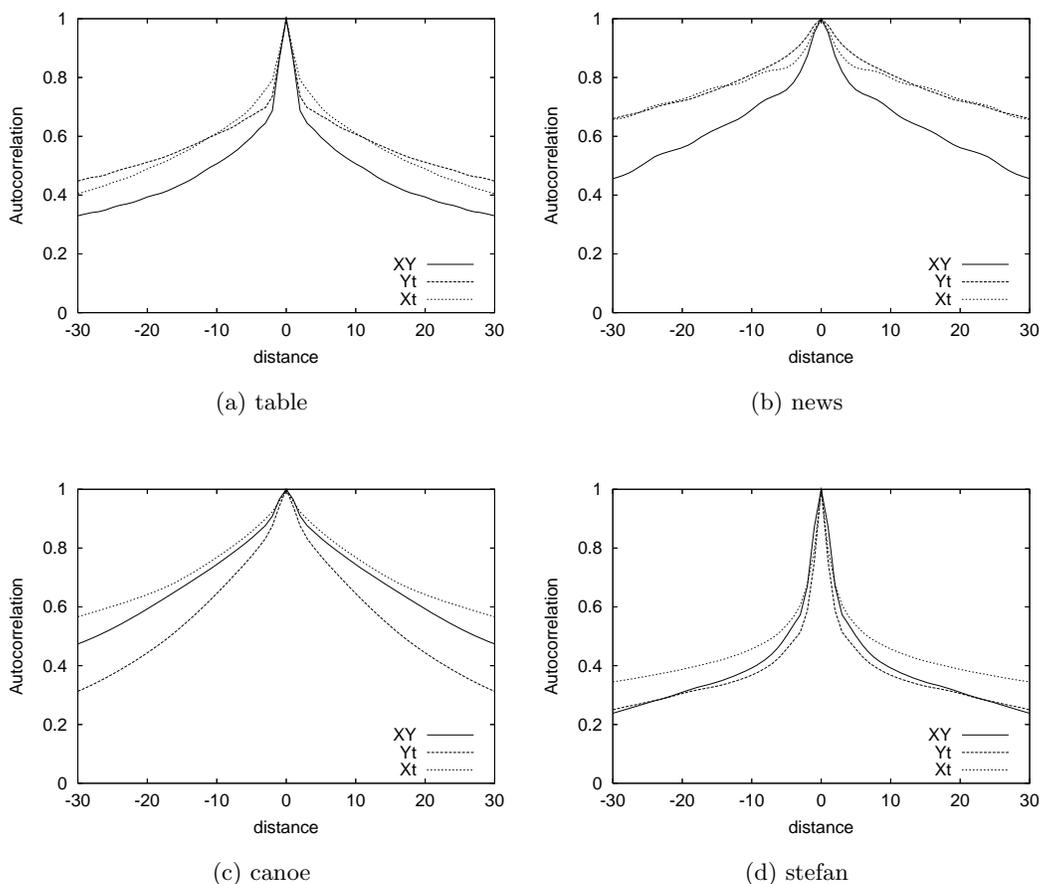


図 4-11: 2次元の自己相関関数

### 4.3 フレーム間予測

これまで未検討であったフレーム間予測符号化を時空間画像に適用し、その符号化効率を調査する．これまで用いていた JPEG 2000 はイントラ符号化が行われているため、時空間画像を取得する際に選択されなかった空間の残りの1軸の相関は利用されていない．フレーム間予測を行う符号化方式を時空間画像に適用し、空間画像の結果と比較することで、これまでと異なる傾向が得られることが考えられる．

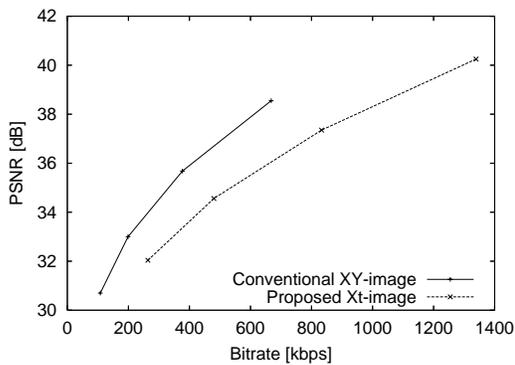
フレーム間予測を用いた代表的な符号化方式である H.264 を用い、フレーム間予測が行われた際の空間画像と時空間画像の符号化効率の比較を図 4-12 に示す．H.264 の参照ソフトウェアである JM を用いて実験を行い、このときの実験条件を表 4-2 に示す．

図 3-2 に示された table, news のように、JPEG 2000 による符号化では効率が改善されていたシーケンスであっても、H.264 では改善されないことが確認される．いずれのシーケンスにおいても、高ビットレートになるにつれ、時系列上での符号化効率の方が、時空間画像の符号化効率よりも大きく改善されている．

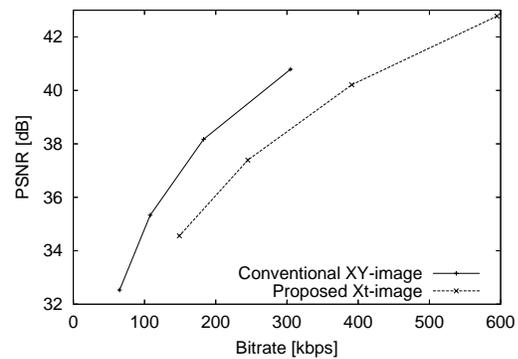
この結果からフレーム間予測を用いた符号化では、通常的时间に沿ったフレーム間で予測

表 4-2: 実験条件

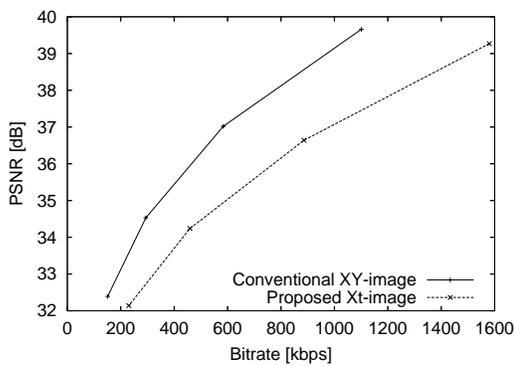
	XY	Xt
Resolution [pels]	352x288	352x300
Number of Frames	300	288
Frame rate [fps]	30.0	28.8
JM Version	10.2	
Profile	High Profile	
QP	I, P= 24, 28, 32, 36(fixed)	
GOP structure	IPPPP	
Number of Reference Frames	1	
MC Block Size	16x16 ~ 4x4	
Search Range	16	
RD Optimization	on	



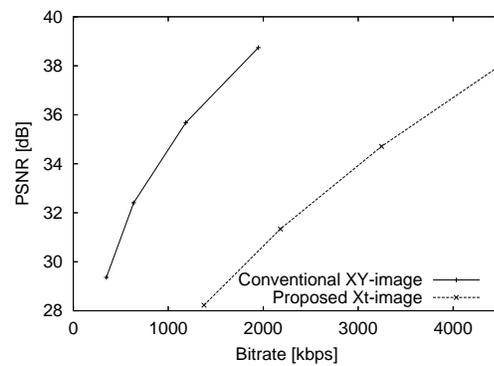
(a) table



(b) news



(c) canoe



(d) stefan

図 4-12: レート対歪み特性

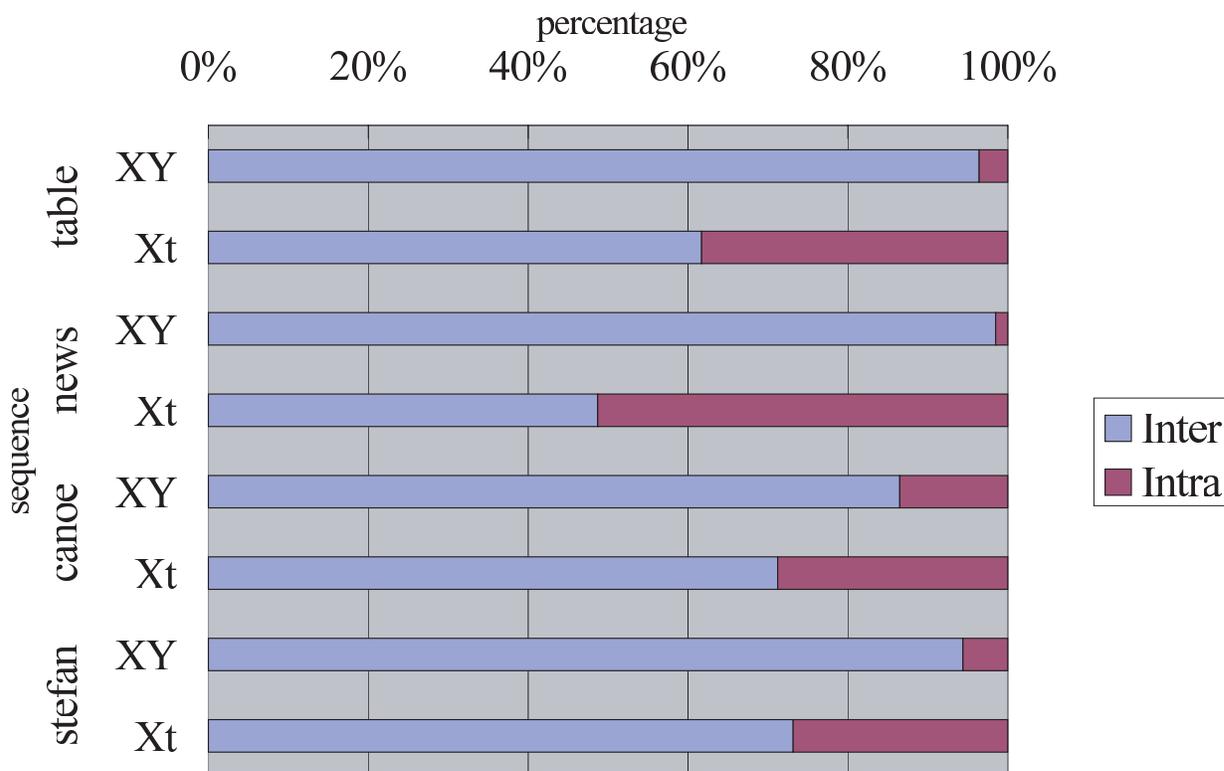


図 4-13: Inter/Intra モード利用割合

符号化を行う方が圧倒的に効率が良いことが示される．このことについて考察を行う．

IピクチャとPピクチャにおける利用モード数を表 4-3 に示す．また，それぞれの割合を図 4-13 に示す．

いずれの画像においても，空間画像に比べ時空間画像の方が Intra 予測が多く選ばれていることが確認される．一般に動画像信号は時間の相関が強いため，この相関を利用したのがフレーム間動き補償予測である．空間画像では，フレーム間の相関が空間相関に比べて強いいため，フレーム間予測が多く選択されているが，逆に時空間画像ではフレーム内での相関の方が，フレーム間の相関よりも強いため．フレーム内での動き補償予測が多く選択されている．カメラの動きの少ない news, table といったシーケンスでは，それがより顕著に表れている．canoe, stefan は前者 2 つほどではないが，時空間画像での P ピクチャが空間画像のそれを上回りはしない．

通常の動画像では，時間方向にフレーム間予測が適用される時間の相関が強いため，提案手法では，フレーム作成時に選択されなかった残りの空間軸方向に動きベクトルが発生するが，図 4-10 に示されたように時間軸の相関が強い table, news などのシーケンスでは Intra 予測が多く利用されている．

また，JPEG 2000 を用いた時空間画像符号化では PSNR が周期的に発生する問題があり，4.2.1 節で原因，解決策を検討した．Wavelet 変換によってサブバンド分割された係数がレート制御によって損失された結果，合成時に Wavelet 関数の基底に倣って符号化損失が生じ，

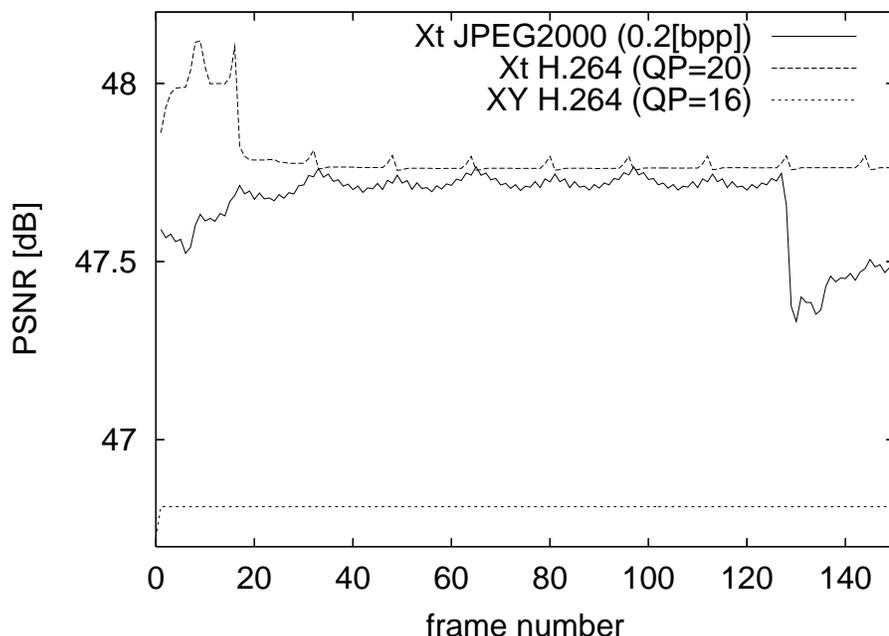


図 4-14: 時間的変動なし時空間画像 PSNR

時間方向に見た場合に周期的な損失として観測されると考察した．この問題は Wavelet 変換によるものであり，DCT 変換を用いた H.264 では生じないことを確認する．符号化するシーケンスには周期的な PSNR の変動がより顕著に現れていた，ある 1 フレームが連続するシーケンスを用いた．JM を用いて時空間画像を符号化し，復号後の時間に沿ったフレーム毎の PSNR を比較した結果を図 4-14 に示す．

図 4-14 より，時空間画像の JPEG 2000 による符号化では 1 フレームずつ PSNR が細かく変動しているが，H.264 による符号化では発生していない．このことから，前述した問題が時空間画像によるものではなく，JPEG 2000 によるものであったと言える．また，H.264 による符号化結果では 16 フレームおきに規則的に PSNR が上昇していることが確認される．これは 16x16 画素のマクロブロックの境界で発生している歪みが蓄積された結果と考えられる．また JPEG 2000 による符号化結果でも 16，ないしは 32 フレーム周期で規則的に変化していることが確認されるが，これは低サブバンド信号までビット切捨てが及んだことによって生じているものであり，コードブロックのサイズによるものではない．なお，同一の空間画像が並んだフレーム間での予測符号化結果は，当然のように，2 フレーム目以降一定となる．また，時空間画像の H.264 符号化における時空間ブロックの符号量は，空間画像のそれより，遥かに大きくなる．これには，通常の時間に沿った符号化では，2 フレーム目以降 Intra ブロックが発生しないのに比べ，時空間画像では常に Intra 予測が行われうるためである．

入力されたシーケンスに対して最適な符号化が行われているだけであるので，時空間画像を時間の並びに再構築した際には，傾向が変わってくることも追記しておく．

#### 4.4 おわりに

本章では，時空間画像の符号化により生じる影響について検討を行い，改善を試みた．まず，時空間画像の JPEG 2000 による符号化時に発生する問題について検討した．また，符号化効率の改善に寄与する要素について検討を行った．最後に，フレーム間予測を用いた時空間画像の符号化効率について検討した．

表 4-3: モード利用割合 (300[kbps])

		table		news	
		Conv. XY	Proposed Xt	Conv. XY	Proposed Xt
I Picture	Intra 4x4	40	73	114	1
	Intra 8x8	230	79	187	202
	Intra 16x16	126	266	95	215
P Picture	Copy	73598	41829	82414	33292
	16x16	19483	12909	13571	11395
	16x8	5830	6043	4102	4232
	8x16	5627	4792	5276	4717
	8x8	9614	8436	11231	4797
	Intra 4x4	529	6121	929	33296
	Intra 8x8	1999	13661	598	8276
Intra 16x16	1724	26175	283	19961	
		canoe		stefan	
		Conv. XY	Proposed Xt	Conv. XY	Proposed Xt
I Picture	Intra 4x4	50	15	144	2
	Intra 8x8	199	228	230	70
	Intra 16x16	147	175	22	346
P Picture	Copy	52310	46674	47551	43079
	16x16	30094	21039	28292	23843
	16x8	8722	8614	11239	11814
	8x16	5652	4979	9499	5396
	8x8	5621	4125	15190	3623
	Intra 4x4	662	1570	1313	1977
	Intra 8x8	11061	25367	3699	20716
Intra 16x16	4282	7598	1621	9518	

## 第5章

# 高フレームレート映像の時間解像度変換

### 5.1 はじめに

本章では、高速度カメラで撮影された高フレームレート映像に対して、時系列信号のフィルタリング処理により時間解像度変換を行う。まず、高フレームレート映像の特徴について説明する。次に、時空間画像のアプローチにより、高フレームレート映像に対する符号化について検討を行う。また、フィルタリングにより階層的な符号化を行うことで、高フレームレート映像の時間解像度変換について検討を行う。

### 5.2 高フレームレート映像

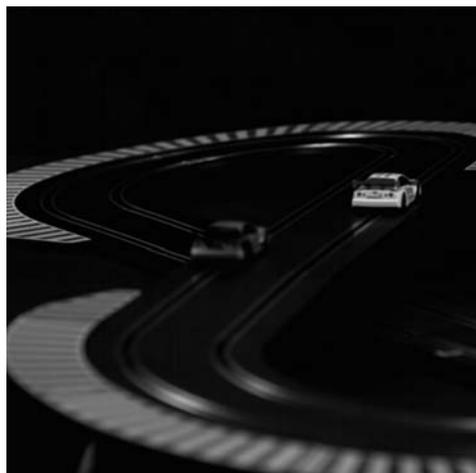
高フレームレート映像の特徴について述べる。

本章の検討に用いるシーケンスの、それぞれ代表的なフレームを図 5-1 に示す。なお、この映像ソースは、NAC 社製デジタルハイスピードカメラ (MEMREACAM fx RX-3) を用いて撮影されたシーケンスである。撮影された映像の空間解像度は 1280x1024、撮影フレームレートは 1000[frame/s] であり、シャッター時間はフレーム間隔に等しく、開口率は 100[%] である。なお、ここでいう開口率は、フレーム間隔 [s] に対するシャッターが開いている時間 [s] の割合である。

いずれのシーケンスもカメラの位置、パン、ズームなどは固定されており、オブジェクトの動きによってのみ瞬間的な動きが発生し、それが画像データとして得られている。いずれもオブジェクトの位置、及び形状の変化が高速に起こり、時間的に高密度にオブジェクトの変化を確認できるシーケンスである。

まず、自己相関関数を求める。第 4 章と同様に空間の 2 軸と時間軸のそれぞれの系列について自己相関関数の平均を図 5-2 に示す。なお、それぞれのシーケンスの画像は、各空間画像における両端を落とし、1024x1024[pels] のものを用意し、1024[frame] 並べるものとする。これにより、3 次元のブロックを立方体と捉えることが可能となり、それぞれの軸の長さに依存しない比較が可能となる。

ここで述べた高フレームレート映像は、時間方向に非常に密であり、実際のオブジェクトの動きをより正確に捉えていると言える。今後はこのシーケンスを元に解像度変換を行う。



(a) car



(b) tennis



(c) object



(d) baseball

図 5-1: 高フレームレート映像サンプル

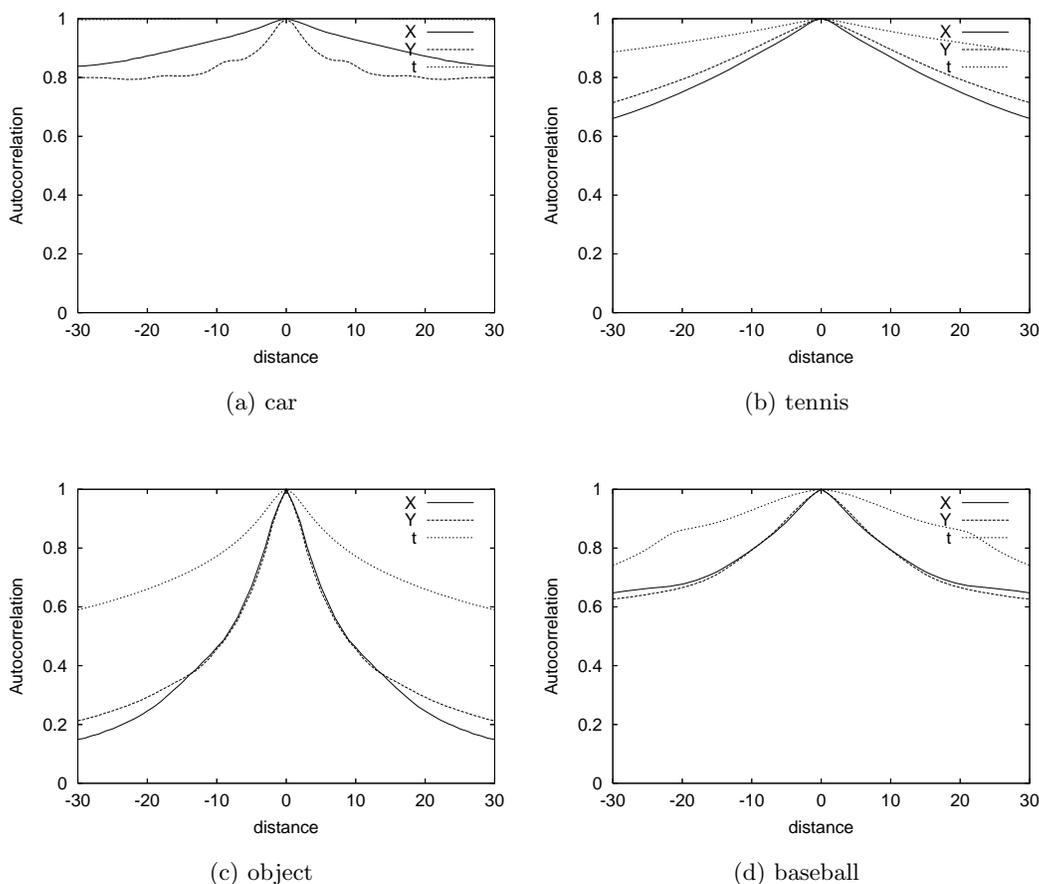


図 5-2: 自己相関関数

### 5.3 符号化効率

高フレームレート映像に対する符号化について検討を行う。

第4章で、時空間画像の符号化について検討を行った。本節では、高フレームレート映像の符号化に対して、時空間画像によるアプローチが有効であることを示す。

まず、フレームレートと符号量との関係について示す。ここでは、目的のフレームレートに合わせてフレームを間引くことで任意のフレームレートの時空間画像を取得し、符号量を比較する。

単位時間を1[s]とし、その時間中における時間解像度を変化させる。空間解像度は $1024 \times 1024$ [pels]、フレーム数はフレームレートに応じて変化させ、500[fps]では500[frame]が並ぶようなシーケンスを符号化対象に用いる。すなわち、実時間は同じ1[s]でありながら、時系列においてサンプル数の異なる信号が並んだシーケンスの符号量の比較を行う。なお、時間解像度のダウンサンプリングには単純なフレーム間引きを行った。時間相関を考慮した符号化を行うために、第4章で検討した時空間画像の符号化を行った。JPEG 2000によるフレーム内符号化を行い、 $5/3$ 可逆フィルタ、及び $9/7$ 非可逆フィルタを用いたいずれもフルビットでの処

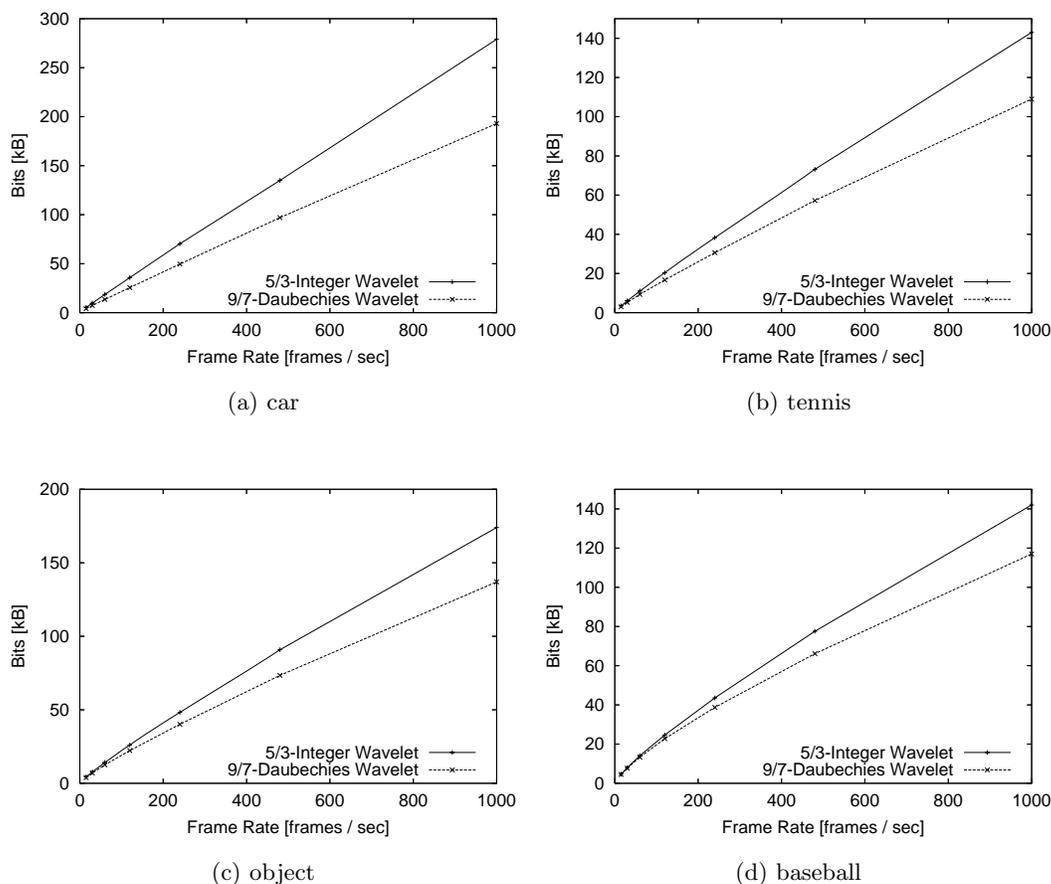


図 5-3: フレームレートと符号量の関係

理により、両者の符号量を比較した。フレームレートと符号量の関係を図 5-3 に示す。

いずれのシーケンスも、程度の差こそあれ、上に凸のグラフとなっている。このことから、フレームレートの増加に伴い、符号化対象となるサンプル数が比例して増加しても、符号量はそれに従わないことがわかる。とりわけ、9/7 非可逆フィルタでは、フレームレートの増加に対する圧縮の効果が大きい。シーケンスごとに見ると、(a)car の 5/3 可逆フィルタによる結果は、特にフレームレートの増加から受ける効果が弱く、ほぼ比例的に符号量が増加してしまっている。この映像シーケンスは奥行きがあり、空間内でぼけた領域が比較的多いため、フレームレートの増加による時系列信号の相関に対する影響が弱いことが考えられる。

また、5/3 フィルタによる符号化は可逆圧縮であったが、9/7 フィルタは非可逆圧縮であるため、フルビットであっても劣化が生じる。そして、その欠落する部分は高周波成分であり、高速度撮影によって発生した雑音が適宜抑えられ、より符号化に適した信号が得られると考えられる。

また、いずれのシーケンスも空間画像の符号化に比べ、符号化効率が高いことが確認されている。空間画像のフレーム内符号化では、フレームレートの高低に関わらず、フレーム数

に応じてほぼ一定の符号量を割く必要がある。第4章の考察より、本章で実験に用いているような、カメラが固定され背景が変化しないような画像に対しては、時空間画像のアプローチによる符号化が有効であることが示されている。

なお、検討に用いたシーケンスのオリジナルの開口率は100[%]であっても、サンプルを間引いた後の開口率はそのフレームレートによって変化する。たとえば、1000[fps]から250[fps]に間引いた場合、フレーム間隔に対するシャッタ開口時間の割合は25[%]となる。同じフレームレートで比較した場合、開口率の低い方が画像の鮮鋭度が高くなり、時系列上での滑らかさは失われる。逆に、開口率が高いほど、空間的にぼやけて見えるが、時系列上では滑らかな動きが再現される、時空間画像としての符号化にも有利となる。撮像系の積分効果をモデル化したレート変換を行うことで、より高品質な映像の再現が可能になると考えられる。

#### 5.4 時間解像度変換

本節では、これまで検討を行ってきた時空間画像の符号化のアプローチを利用した時間解像度変換法について検討する。これまでの考察により、すでに高フレームレート映像を時空間画像として符号化した際の有効性は示されている。時系列信号が画像の1軸として符号化されることにより、高フレームレート映像における時系列信号を符号化することの有効性が確認された。

この提案手法では、JPEG 2000 を用いたことによるサブバンド分割が時系列信号に対して行われる。低域成分が再帰的に分割されることで、階層的なフィルタリングが施されている。ある解像度レベルにおいて、低域成分のみから復号を行うことにより、その解像度レベルにおける高周波成分の失われた信号が得られる。前節では、フレームレート変換を間引きにより実現したが、本手法によりローパスフィルタがかかった信号が得られ、フレーム間での滑らかさが得られることが期待される。

本節でも、これまで同様 JPEG 2000 によって時系列信号を符号化し、所望の解像度変換を行う。JPEG 2000 では、5/3 Integer Wavelet と 9/7 Daubechies Wavelet が定義されているが、タップ数の長さによって、より多くの時系列信号に影響を与えてしまう恐れがある。タップ数の短い可逆整数型フィルタを選択することで、より時系列の並びにおける前後の信号に特化したフィルタリングを行う。なお、任意の解像度レベルでのサブバンド信号を切り捨てて解像度変換を行うため、可逆圧縮については考慮しない。

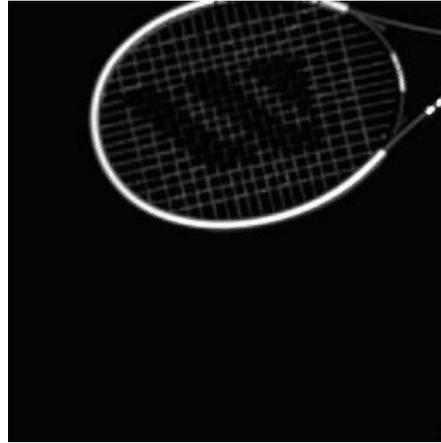
3次元信号中における、時系列信号にのみサブバンド分割を行い、時系列信号に特化した解像度変換を行う。それぞれのシーケンスで、任意のタイミングにおける時系列上の解像度レベルがそれぞれ0,2,4の画像から復号したものを図5-4~5-7に示す。解像度レベル0は原画像である。

いずれも解像度レベルが高くなるにつれ、画像のボヤケが多くなっていることが確認される。これは空間に対して処理が行われたのではなく、いずれも時系列上でローパスがかかったことにより、前後のフレームの情報を含んだボヤケとして空間画像に現れている。

前節で触れた開口率の観点から考察を行う。前節でのフレーム間引きでは、間引き後に



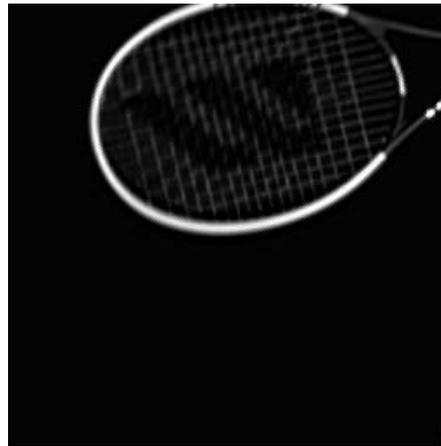
(a) Level = 0



(a) Level = 0



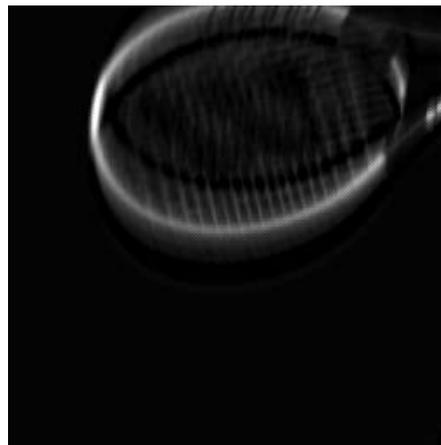
(b) Level = 2



(b) Level = 2



(c) Level = 4



(c) Level = 4

図 5-4: car

図 5-5: tennis



(a) Level = 0



(b) Level = 2



(c) Level = 4

図 5-6: object



(a) Level = 0



(b) Level = 2



(c) Level = 4

図 5-7: baseball

残ったフレームのシャッター開口時間のみが，開口率に関わっていた．しかし，5タップの長さをもつ Wavelet フィルタにより，単純積分ではない低域通過成分が得られる．サンプリング間隔に応じて寄与率が異なるため，シャッタ機能に対するアパーチャ効果を考慮した信号といえる．こうして得られた信号により，空間においては鮮鋭度が失われているものの，時系列上において，物体の動きはより滑らかに得られることを確認した．

再帰的なフィルタリングを行う時空間画像の符号化を，時間解像度変換に応用したことにより，高フレームレート映像からの低域通過信号の表現を可能にした．一方で，解像度レベルに応じて，タップ数により影響を受ける時間幅がより長くなるため，ローパスが強くかかりすぎることが懸念される．視覚特性を考慮し，定性的な解析が必要である．また，時空間画像符号化のアプローチによる， $1/2^n$  に限定されない任意の解像度に変換する手法の検討を行う必要がある．

## 5.5 おわりに

本章では，高フレームレート映像に対する時間解像度変換手法について検討を行った．まず，時空間画像のアプローチにより，高フレームレート映像に対する符号化について検討を行った．また，フィルタリングにより階層的な符号化を用いた高フレームレート映像の時間解像度変換の検討を行った．

## 第6章

### 結論

本章では本論文を総括し、今後の検討課題を述べる。

#### 6.1 総括

本研究では、時系列信号としての映像に着目し、ウェーブレット解析により、動画像の高フレームレート化に対する基礎検討を行った。時系列信号を時空間画像における画像の1軸としてとらえ、フレーム相関を直接的に利用する符号化手法について検討を行った。その手法により、カメラ固定で撮影されたシーケンスに対して、効率的な符号化を実現した。また、時系列信号をウェーブレットフィルタによって階層的に符号化することにより、高フレームレートからの時間解像度変換を行う手法についても検討を行った。同手法を用いた時間解像度変換により、 $1/2^n$  のフレームレートにおける高品質な動画像表現が可能となった。

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、本研究の要素技術となる JPEG 2000 について説明した。また、動画像符号化方式として本研究で検討を行う、Motion-JPEG 2000 及び H.264/AVC について述べた。最後に検討課題とする解像度変換の従来手法について説明した。

第3章では、本研究で着目した動画像における時系列信号について述べた。まず時空間画像の概要について述べ、次にその特徴が用いられている関連研究について説明した。次に時空間画像の符号化のアプローチについて説明した。最後に時系列信号の観点から、高フレームレート映像について述べた。

第4章では、第3章で述べた時空間画像の符号化の結果について考察し、改善を試みた。まず、時空間画像の JPEG 2000 による符号化時に発生する問題について検討した。また、符号化効率の改善に寄与する要素について検討を行った。最後に、フレーム間予測を用いた時空間画像の符号化効率について検討した。

第5章では、高フレームレート映像に対する時間解像度変換手法について検討を行った。まず、時空間画像のアプローチにより、高フレームレート映像に対する符号化について検討を行った。また、フィルタリングにより階層的な符号化を用いた高フレームレート映像の時間解像度変換の検討を行った。

第6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論じている。

## 6.2 今後の課題

今後検討すべき課題を以下に示す．時系列信号を符号化に適用した際の有効性は示されたが，シーケンスによって，時系列上の相関の低い場合がある．そういったシーケンスに対して，適応的に有効な軸を選ぶことによる符号化効率の改善について検討を行う必要がある．また，2次元の JPEG 2000 で行われている処理を 3次元で行うことにより，時空間ブロックをより効率的に圧縮できることが考えられる．また，それにより，空間解像度変換も，同等の処理で行うことができると考えられる．解像度変換において， $1/2^n$  に限られない，任意の解像度に変換する手法について検討する必要がある．また，その定性的な解析についても検討の余地がある．

## 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。

貴重な御意見，様々な御提案を頂いた渡辺研 AVS ゼミの皆様に御礼申し上げます。

惜しみなく知識・助言を提供して頂いた石川 孝明氏に心から感謝致します。

常に友好的に接して下さった河村 圭氏に心から感謝致します。

学部時代から渡辺研同期として，様々な助言を与えてくれた，伊谷 裕介氏，後藤 崇行氏，山本 勇樹氏に心から感謝いたします。

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 19 年 3 月 15 日

## 参考文献

- (1) <http://www.youtube.com/>.
- (2) “カメラの枠を吹き飛ばす ソニーの超高速 CMOS センサ,” 日経エレクトロニクス, 2006.02.27, 2006.
- (3) 竹間忠夫, “夢のスーパーハイビジョンに挑む,” 日本放送出版協会, 2005 .
- (4) “超高精細化,” 日経マイクロデバイス, 2006.11, 2006 .
- (5) 吹抜敬彦, “テレビ信号の見え方の説明に仮現運動は要るか,” 映像情報メディア学会誌, Vol. 60, No. 12, pp. 2015-2018, December 2006 .
- (6) 坂東幸浩, 高村誠之, 上倉一人, 八島由幸, “高フレームレート映像信号に対する符号化特性解析,” 情報処理学会 研究報告 2005-AVM-48, pp . 13-17, March 2005.
- (7) ISO/IEC JTC1/SC29 WG1, “ISO/IEC 15444-1 JPEG2000 Part 1:Core coding system,” July 2002.
- (8) <http://www.jpeg.org/jpeg2000>.
- (9) 福原隆浩, 次世代動画像符号化方式 Motion-JPEG2000, トリケップス, 2002.
- (10) ISO/IEC JTC1/SC29 WG1 N2250, “Motion JPEG 2000 Final Draft International Standard 1.0,” September 2001.
- (11) Ingrid Daubechies, “Ten Lectures on Wavelets,” Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- (12) 小野定康, 鈴木純司, わかりやすいJPEG2000 の技術, オーム社, May 2003.
- (13) 貴家仁志, “JPEG2000 を中心とした画像圧縮技術の新しい流れ,” CQ 出版社, インターフェース, pp.46-58, January 2002
- (14) 貴家仁志, 渡邊修, “JPEG2000 符号化アルゴリズムの要素技術,” CQ 出版社, インターフェース, pp.59-71, January 2002.
- (15) <http://www.microsoft.com/whdc/xps/wmphoto.mspx>.
- (16) 木村青司, 荒木淳哉, 福原隆浩, “Motion JPEG2000 における Visual weighting を用いた画質改善の検討,” 画像符号化シンポジウム資料, Vol.16th, pp.75-76, 2001.
- (17) Digital Cinema Initiatives. LLC, “Digital Cinema System Specification v1.0,” July 2005.
- (18) “デジタル・シネマ -HDTV を超える映像システムの登場,” 日経エレクトロニクス 2005.08.15, 2005 .

- (19) 小野定康, 村上篤道, 浅井光太郎, 動画像の高能率符号化 MPEG-4 と H.264 , オーム社, 2005.
- (20) C.Kim, Q.Li, C.-C.J.Kuo, "Fast Intra-Prediction Model Selection for H.264 Codec," Proceeding of the SPIE, Volume 5241, pp.99-110, November 2003.
- (21) 尾知博, シミュレーションで学ぶデジタル信号処理, CQ 出版社, 2001.
- (22) Window of visibility: a psychophysical theory
- (23) 境田慎一, 苗村昌秀, 金次保明, "背景差分法と時空間 watershed による領域成長法を併用した動画像オブジェクトの抽出," 電子情報通信学会論文誌, D-II, Vol.J84-D-II, No.12, pp.2541-2555, December 2001.
- (24) ISO/IEC 14496-2, "Information technology-Coding of audio-visual objects"(2000).
- (25) J.Wang, and E.Adelson, "Representing Moving Images with Layers," IEEE Trans. on Image Processing, vol.3, no.5, pp.625-638, September 1994.
- (26) Y.Wexler, E.Shechtman, and M.Irani, "Space-Time Video Completion," Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004.
- (27) 庄司竜太, 新田直子, 馬場口登, "テロップ消去のための時空間特徴を用いた映像修復," 映像情報メディア学会誌, vol.61, no.1, pp.91-97, January 2007.
- (28) 谷口博康, 関明伸, 古澤春樹, 黒田伸一, 池端重樹, "時空間画像を用いた動画像処理手法の提案 - DTT 法 - ," 電子情報通信学会論文誌, D-II, Vol.J77-D-II, No.10, pp.2019-2026, October 1994.
- (29) 本多充知, 加藤博一, 西田正吾, "時空間画像を用いたカーナビのための自動位置補正," 電子情報通信学会技術研究報告, MVE2006-44, pp.7-12, September 2006.
- (30) 阿久津明人, 外村佳伸, "投影法を用いた映像の解析方法と映像ハンドリングへの応用," 電子情報通信学会論文誌, D-II, Vol.J79-D-II, No.5, pp.675-686, May 1996.
- (31) 岡本啓嗣, 八杉将伸, 馬場口登, "固定長の時空間画像に基づく映像シーンのクラスタリング," 電子情報通信学会論文誌, D-II, Vol.J86-D-II, No.6, pp.877-885, June 2003.
- (32) 加藤幸一, 渡辺裕, 安田靖彦, "時空間画像の符号化に関する一考察," 情報処理学会全国大会, 3W-5, March 2005.
- (33) "KaKadu software," <http://www.kakadusoftware.com/>.
- (34) <http://analysis.nacinc.jp/products/uhscon200/index.html>.

- (35) 吹抜敬彦, 北村彰規 “テレビ撮像系や表示系におけるシャッタ機能の時空間標本化からの考察” 映像情報メディア学会誌, Vol. 59, No. 12, pp. 1841-1846, December 2005 .

## 図一覧

2-1	JPEG2000 符号化アルゴリズム	6
2-2	2分割フィルタバンク	8
2-3	2ステージ分解	8
2-4	H.264/AVC 符号化アルゴリズム	16
2-5	動き補償時間方向フィルタ	20
3-1	時空間画像	22
3-2	時空間画像符号化による平均 PSNR	24
3-3	Hall Monitor (bit rate: 0.5[bpp])	25
3-4	時間的変動なし時空間画像 PSNR	26
4-1	時間的変動なし時空間画像	28
4-2	時間的変動なし時空間画像 Wavelet 係数	28
4-3	非可逆変換の差分画像 (時間的変動なし時空間画像)	29
4-4	提案手法 (ジグザグスキャン)	30
4-5	Hall Monitor PSNR	31
4-6	Table PSNR	31
4-7	canoe PSNR	32
4-8	コードブロックサイズ	33
4-9	コードブロックサイズ (Lossless)	34
4-10	1次元の自己相関関数	35
4-11	2次元の自己相関関数	36
4-12	レート対歪み特性	37
4-13	Inter/Intra モード利用割合	38
4-14	時間的変動なし時空間画像 PSNR	39
5-1	高フレームレート映像サンプル	43
5-2	自己相関関数	44
5-3	フレームレートと符号量の関係	45
5-4	car	47
5-5	tennis	47
5-6	object	48
5-7	baseball	48

## 表一覧

2-1	JPEG 2000 標準方式の各 Part . . . . .	5
2-2	整数型 5/3 フィルタの係数 . . . . .	9
2-3	実数型 9/7 フィルタの係数 . . . . .	9
2-4	デジタルシネマの規格概要 . . . . .	15
2-5	H.264 プロファイル構成 . . . . .	18
4-1	PSNR の平均 / 標準偏差 (Hall Monitor 1.0[bpp]) . . . . .	32
4-2	実験条件 . . . . .	37
4-3	モード利用割合 (300[kbps]) . . . . .	41

## 研究業績

	題 名	発表年月	発表掲載誌	連名者
国内				
(1)	時空間画像の符号化に関する一考察	2005年3月	情処全大	加藤 幸一 渡辺 裕 安田 靖彦
(2)	時空間画像の符号化による時間軸ノイズの低減に関する一考察	2005年3月	信学春季全大	加藤 幸一 渡辺 裕 安田 靖彦
(3)	時空間画像の JPEG 2000 符号化における SNR 揺らぎについて	2005年9月	FIT	加藤 幸一 石川 孝明 渡辺 裕
(4)	時空間画像の JPEG 2000 符号化における SNR の周期的変動に関する考察	2005年11月	PCSJ	加藤 幸一 石川 孝明 渡辺 裕
(5)	時空間画像の JPEG 2000 符号化による PSNR の周期的変動の低減に関する検討	2007年3月	信学春季全大	加藤 幸一 渡辺 裕