

## H.264, MPEG-4 Studio Profile, JPEG 2000 を用いた多バンド動 画像符号化に関する一検討

高村誠之<sup>†</sup> 八島由幸<sup>†</sup>

<sup>†</sup> 〒 239-0847 横須賀市光の丘 1-1 Y-517A

日本電信電話株式会社 NTT サイバースペース研究所

Tel: 046-859-2371 Fax: 046-859-2829

E-mail: {takamura.seishi, yashima.yoshiyuki}@lab.ntt.co.jp

あらまし マルチスペクトル画像は従来衛星画像などでは用いられていたが、近年視覚を対象とする装置であるプリンタ、カメラ、ビデオなどでも次第に用いられつつある。多バンド化することでさまざまな機能が実現される一方、信号の容量はバンド数に比例して増大するため、特に動画像において効率的圧縮が望まれる。本稿では符号化に適したフォーマットを検討し、実際に6バンド映像信号(48 bit/pel)を、JPEG 2000, MPEG-4 Studio Profile および近年標準化された H.264 により符号化した。ビットレート 3.5 bit/pel を境にそれ以上では JPEG 2000 が、それ以下では H.264 が最も高い PSNR を示した。MPEG-4 Studio Profile は常に中間の性能であった。

## A Consideration on Multiband Video Coding with H.264, MPEG-4 Studio Profile and JPEG 2000

Seishi TAKAMURA<sup>†</sup> and Yoshiyuki YASHIMA<sup>†</sup>

<sup>†</sup>NTT Cyber Space Laboratories, NTT Corporation,

Y517A, 1-1 Hikari-no-oka, Yokosuka, Kanagawa, 239-0847 Japan,

phone: +81 46 859 2371, fax: +81 46 859 2829,

E-mail: {takamura.seishi, yashima.yoshiyuki}@lab.ntt.co.jp

Abstract: Multispectral imaging, which has been common in satellite image acquisition to a certain extent, is recently being used in visual imaging systems such as printers, digital/analog cameras, still/video cameras, projectors and even displays. The reason is the variety of functionalities it provides. As the amount of image data is proportional to the number of bands, efficient compression of such data sets, particularly for video data, is necessary. This paper provides a basic investigation of multispectral video coding, such as inter-band decorrelation, visible/invisible separation and non-linear signal transform prior to encoding. We use a six-band video sequence (48 bit/pel) as the test data. The video/image codecs of JPEG 2000, MPEG-4 Studio Profile, as well as the newly developed standard H.264, are compared in a coding experiment. The results show that JPEG 2000 yields higher PSNR than the other codecs at bit rates of around 3.5 bit/pel and above. H.264 performs best at lower bit rates. MPEG-4 Studio Profile is always the second best codec.

# 1 はじめに

マルチスペクトルに基づく映像技術は、気象・海洋・水文などを目的とした衛星リモートセンシングですでに広く用いられており、その効率的な符号化方式の提案もなされている [1]。

テレビジョンやカメラなど、人が見るための画像データにおいては、長らく3バンド入力、3原色表示といった3自由度の装置が用いられてきた。人間の色覚が3自由度しかもっていないことがこの理由として挙げられる。しかしながら、色の本質的情報は無次元のスペクトル空間にこそ存在するのであって3次元表現は大きく情報が削減された射影にすぎない、という認識から近年マルチスペクトルに基づく映像技術が注目を集め、かつ活発に開発されている [2, 3]。このようなシステムが扱うデータは、そのバンド数に比例して大容量となる。そこでこの効率的な符号化方式が望まれる。

本稿文では、多バンド映像技術の性質について述べた後、色覚特性を考慮した信号変換方式および主成分変換に基づいた信号変換方式を提案する。ついで異なる符号化方式を用い、符号量-歪特性を信号変換方式別に測定した。用いた符号化方式は多バンド符号化が可能な MPEG-4 Studio Profile [4] および JPEG 2000 [5]、近年標準化された H.264 [6] である。

## 1.1 多バンド映像の性質

無限次元のスペクトル空間において、3自由度を持つ人間の色覚空間は、CIE 等色関数と呼ばれる3つの基底  $x, y, z$  により張られる。通常はこの内部が目で見える空間となる。あるスペクトルと等色関数との内積がその3測色値  $X, Y, Z$  である。3測色値の代わりにスペクトルを用いることで、例えば正確な色の取得・再現、照明に依存しない色再現、等色関数の個人差の吸収といった機能性が実現される。

## 1.2 従来方式

色覚が3自由度しか持たないことから、スペクトル情報を目に見える「可視バンド」と見えない「不可視バンド」に分離することで、スペクトルに基づく機能性と、従来システムとの互換性の両方を実現できる。新しいタイプのスケーラビリティと考えることもできる。

そのようなデータフォーマットの提案として Keusen [7] のものがあるが、符号化効率化を考慮したものではない。

礎 [8] は多バンド信号を対数変換後、主成分変換を施す方法を提案しているが、非線形変換である対数変換が前段に入るため、得られる値は測色値との対応が直接つかず、可視・不可視分離といった色覚的な特性を反映させることができない。

また、マルチスペクトル動画の符号化効率化に関する研究はわずかである。高村 [9] は可視・不可視バンドを分離し、それぞれに主成分変換を施す方法を提案した。この方法は前述のスケーラビリティを実現するものであるが、電力集中度の観点からは、可視・不可視分離による効率低下が懸念される。

そこで本稿では、この可視不可視分離変換方式と、全バンドを通して主成分変換を行った場合との、符号化後の符号量-歪性能の違いを考察する。また、歪は、上記報告 [9] では変換後空間において測定されていたのに対し、本稿では復号画像に上記変換の逆変換を施した、原信号と同じ多バンドの空間において測定した。

# 2 多バンド映像信号処理

本章では、容量の多さから高能率符号化が望まれる多バンド映像信号の効率的な信号変換方式を提案する。以後、バンド数を  $N$  とする。

## 2.1 主成分変換行列 $T$

従来手法として、バンド間主成分変換を考え、その  $N \times N$  変換行列を  $T$  とする。変換後の各成分の映像は、符号化器に応じた空間解像度に適宜変換される。

## 2.2 可視・不可視分離行列 $V$

$N$  次元の多バンド信号を3次元の可視信号と  $N-3$  次元の不可視信号に分離することを考える。

まず、 $x, y, z$  をそれぞれ等色関数相当の  $N$  次元列ベクトルとする。これらは入力装置のバンド別分光感度に依存していることに注意する。任意の多バンド信号 (列ベクトル)  $a$  から、3測色値  $X, Y, Z$  を  $X = x^t a$  のように求めることができる。

Gram-Schmidt 直交化により以下のように直交基底

$v_x, v_y, v_z$  を展開する:

$$v_x := x/\|x\| \quad (1)$$

$$\mathbf{P} := \mathbf{I} - \mathbf{x}\mathbf{x}^t \quad (2)$$

$$v_y := \mathbf{P}\mathbf{y}/\|\mathbf{P}\mathbf{y}\| \quad (3)$$

$$\mathbf{P} := \mathbf{P}(\mathbf{I} - v_y v_y^t) \quad (4)$$

$$v_z := \mathbf{P}\mathbf{z}/\|\mathbf{P}\mathbf{z}\| \quad (5)$$

$$\mathbf{P} := \mathbf{P}(\mathbf{I} - v_z v_z^t) \quad (6)$$

ここで  $\mathbf{I}$  は  $N \times N$  単位行列である。任意の多バンド信号  $a$  は、式 6 の射影行列  $\mathbf{P}$  を施されると  $x, y, z$  の直交補空間へ写像される。

残る  $N - 3$  次元の直交基底  $v_i$  ( $i = 4 \dots N$ ) は、例えば以下のようにして求めることができる。

$$\text{for } (i = 4; i \leq N; i++) \{$$

$$v_i := \mathbf{P}\mathbf{r}_i/\|\mathbf{P}\mathbf{r}_i\| \quad (7)$$

$$\mathbf{P} := \mathbf{P}(\mathbf{I} - v_i v_i^t) \quad (8)$$

}

ここで  $\mathbf{r}_i$  は、 $v_x, v_y, v_z, v_j$  ( $j < i$ ) と一次独立な任意の非零ベクトルである。 $v_i$  は人間の色覚には全く関与しないがマルチスペクトル情報を構成する本質的要素となっている。

式 2 における  $\mathbf{P}$  の初期ランクは  $N - 1$  であり、 $\mathbf{P}$  が更新される度に (式 4, 6, 8)、ランクは 1 ずつ減少し、最終的に 0、つまり  $\mathbf{P}$  は零行列となる。

次いで、可視・不可視信号それぞれ別個に主成分変換を施す。その変換行列をそれぞれ  $\mathbf{P}_v$  および  $\mathbf{P}_i$  とする。サイズはそれぞれ  $3 \times 3$  および  $(N - 3) \times (N - 3)$  である。これらは  $v_n$  と異なりシーケンス・被写体依存である。

これらを結合した順変換行列  $\mathbf{V}$  (正規直交である) は以下ようになる:

$$\mathbf{V} = \begin{bmatrix} \mathbf{P}_v & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_i \end{bmatrix} [v_x, v_y, v_z, v_4, v_5, \dots, v_N]^t \quad (9)$$

変換後の各成分は、用いる符号化器の色差フォーマットに応じて空間解像度を変換する。

## 2.3 提案信号フォーマット

H.264 は YUV 4:2:0 以外のフォーマットをサポートしないため入力信号をこれに沿うよう変換する。本稿で扱う入力信号は、3.1 節にて述べるように 6 バンド信号であるので、これを二本の 4:2:0 画像シーケンスに分離し別個に符号化する。

行列  $\mathbf{T}$  を用いる場合は、第 1、第 2 主成分を Y にあたる最も解像度の高い信号とし、残る第 3~6 主成分を U, V にあたる空間解像度が縦横 1/2 の信号とする。第 1,5,6 成分および第 2,3,4 成分からそれぞれ疑似 YUV 4:2:0 シーケンスを作成する。このフォーマットを“420420”と呼ぶことにする。

行列  $\mathbf{V}$  を用いる場合は、寄与度の最も高い第 1、第 4 主成分に最も高い解像度を割り当て、残りの第 2,3,5,6 成分の空間解像度を縦横 1/2 とする。次いで可視の第 1,2,3 成分および不可視の第 4,5,6 成分からそれぞれ疑似 YUV 4:2:0 シーケンスを作成する。このフォーマットを“xyz420420”と呼ぶことにする。

全バンドフル解像度に対応する JPEG 2000 と MPEG-4 に対しては、空間解像度低減は行わず、 $\mathbf{T}$  により 6 次元主成分変換されたフォーマットを“444444”と呼び、 $\mathbf{V}$  により可視不可視分離されたフォーマットを“xyz444444”と呼び用いることとする。

## 2.4 変換後信号の非線形処理

本稿で符号化する信号はすべて 8bit デジタル信号としている。この表現は、数値を量子化ステップ 1 で量子化し [0,255] でクリッピングすることに相当する。仮にデジタル化がない場合、 $\mathbf{T}$  または  $\mathbf{V}$  による順変換後、 $\mathbf{T}^{-1}$  または  $\mathbf{V}^{-1}$  により逆変換を行えば元の信号に戻る。また両者は正規直交変換であるため、符号化を伴う場合 PSNR は変換後・変換前いずれの空間で測定しても同一となる。しかしながら、順変換後デジタル化がある場合はその限りではなく、特にクリッピングが大きな誤差を招く場合がある。そこで、本稿では変換後の第 1 主成分 (可視不可視分離の場合はそれぞれの第 1 主成分の、計 2 成分) をその範囲に応じてスケールし、さらにガンマ補正と同様の処理を施した後に 8bit デジタル化する。一連の処理を本稿では「非線形処理」と呼び、これにより劣化を抑制している。

この非線形処理を施した場合と施さない場合の発生誤差の定量検討については 3.3 節において述べる。

## 3 計算機シミュレーション

### 3.1 実験映像データ

ここで用いた映像信号は通信・放送機構ナチュラルビジョンリサーチセンタで開発された 6 バンド HDTV

表 1: 6次元主成分変換 (T) と二つの 3次元主成分変換 (V) による電力集中度の比較

成分	T		V	
	電力 %	累積電力	電力 %	累積電力
1	82.1	82.1	63.0	63.0
2	7.1	89.2	2.4	65.4
3	5.1	94.3	1.0	66.4
4	2.9	97.2	33.0	99.4
5	2.2	99.4	0.4	99.8
6	0.6	100.0	0.2	100.0

カメラにより取得されたものを、720x480 30P 各バンド 8bit(48 bit/pel) に変換した。長さは 30 フレームである。以後これを符号化実験や PSNR 測定の原データとした。

### 3.2 T, V の電力集中度

原データに T と V を施した結果の、各成分への電力分布を表 1 に示す。T では電力が低成分へよく集中していること、V では可視空間に約 2/3(66.4%) の電力が配分されていることがわかる。

### 3.3 非線形処理の効果

図 1 に、T 変換による第 1 主成分に対し非線形処理を施さなかった場合と施した場合のヒストグラムの違いを示す。この右図では、信号がより均一に分布すると同時に飽和も避けられていることが見て取れる。

表 2 に、T 変換の後、非線形処理を介さず 8bit デジタル化し  $T^{-1}$  を施した場合に重畳した雑音を示す。信号の飽和(図 1 左)のため、大きい負の差分が発生している。従って空間解像度は落していないにも関わらず、PSNR は平均 37.16dB と非常に低くなっている。

非線形処理を介した場合の結果を表 3(444444) に示す。PSNR が 58.40dB へと大幅に改善していることがわかる。

### 3.4 空間再標本化損

原信号に対し、順変換・ダウンサンプル・非線形処理・8bit デジタル化・逆非線形処理・アップサン

表 2: 非線形処理がない場合に重畳するデジタル化雑音 (単位 dB) および差分

成分	444444	
	PSNR	差分
1	40.98	-93..1
2	40.13	-107..1
3	38.90	-109..1
4	36.88	-136..1
5	35.82	-152..1
6	34.19	-184..1
平均	37.16	

プル・逆変換を施した結果重畳した雑音を表 3 に示す。444444 フォーマットは殆んど誤差がなく、420420 フォーマットでは約 44 dB となった。すなわち、いかなる符号化方式であっても 420420 フォーマットを使う限りはこの PSNR 値が上限となる。

### 3.5 映像符号化方式

上述の方法で変換した 6 バンド映像シーケンスを MPEG-4 Studio Profile, JPEG 2000, H.264 により符号化した。これらの方式の比較を表 4 にまとめた。

MPEG-4 Studio Profile は通常の 3 バンドに加え、奥行き・視差・アルファなどの情報を最大 3 バンドまで符号化できる。B ピクチャは持たない。本実験での符号化条件を表 5 に示す。

JPEG 2000 は任意の色差解像度、16,384 バンドまでの符号化をサポートしている(表 4)。符号化には JJ2000[10] ソフトウェアを用い、各フレームを個別に符号化した。表 6 に符号化条件を示す。

H.264 符号化においては、JM7.3[11] を用いた。符号化条件を表 7 に示す。

### 3.6 符号量-歪性能

各種符号化・変換方式における、ビットレート (bit/pel) と PSNR(dB) の関係をグラフ化し図 2 に示す。H.264 符号化のビットレートは、二つのビットストリームの和から求めたものである。

約 3.5 bit/pel で全方式の PSNR がほぼ同一値(約 43 dB)となった。またそれ以下のビットレートで H.264 が、それ以上のビットレートで JPEG 2000 が最も高い PSNR を与えた。可視・不可視分離および非分離

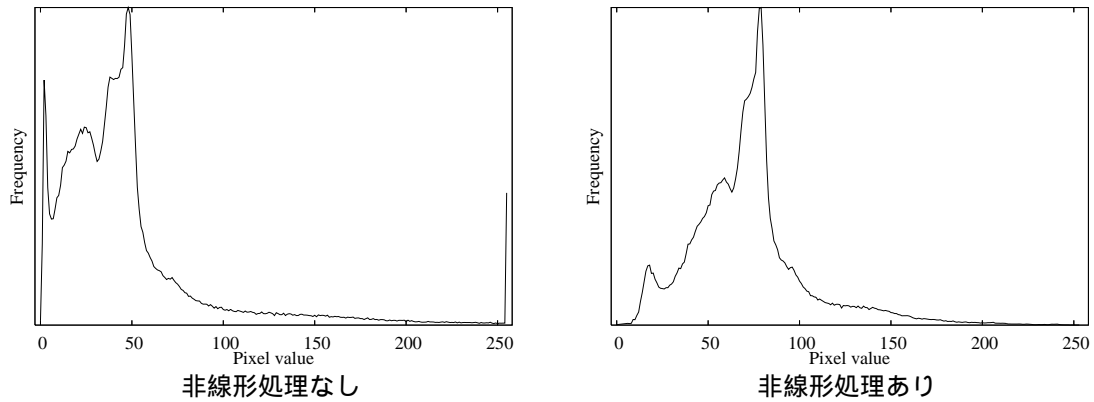


図 1: 第 1 主成分に対する、非線形処理の有無によるヒストグラムの違い

表 3: 非線形処理を行った場合に重畳するデジタル化雑音 (単位 dB) および差分

成分	444444		420420		xyz444444		xyz420420	
	PSNR	差分	PSNR	差分	PSNR	差分	PSNR	diff.
1	58.96	-1..1	44.05	-30..24	58.81	-1..1	41.88	-42..44
2	58.67	-1..1	47.35	-12..19	58.91	-1..2	43.56	-46..35
3	58.48	-1..1	47.56	-32..27	58.58	-1..3	46.60	-20..30
4	58.24	-1..1	41.66	-50..32	58.42	-1..1	43.91	-34..26
5	58.28	-1..1	44.73	-38..22	58.36	-1..1	44.00	-30..36
6	57.84	-2..1	42.58	-33..45	62.73	-1..1	56.04	-7..7
平均	58.40		44.12		59.08		44.48	

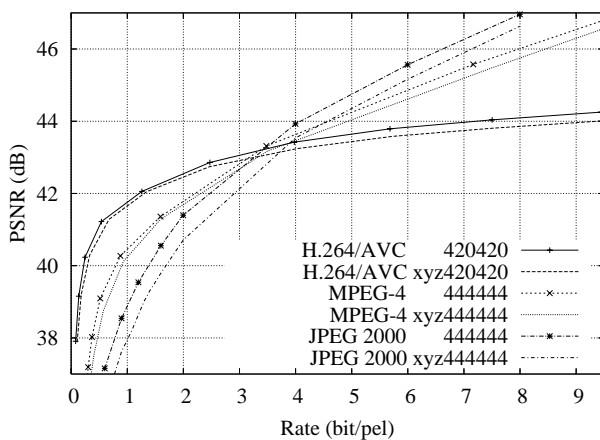


図 2: 符号化効率の比較

(たとえば xyz444444 と xyz444444) の違いは、H.264 と MPEG-4 で約 0.4 dB であり、JPEG 2000 では 1.0 dB 弱であった。

## 4 おわりに

本稿では、多バンド映像信号の性質を概説した後、符号化を指向し色覚に基づいた信号フォーマットを提案した。信号の線形変換後、デジタル化前に非線形処理を施すことが雑音抑制に効果的であることがわかった。

ついで 3 つの符号化方式により符号化を行い定量比較した。全バンド主成分変換に対する可視・不可視分離方式の性能低下に関しては、H.264 と MPEG-4 Studio Profile においては低下はほぼ無視できる程度であることがわかった。

各符号化方式において最適な符号化パラメータやバンド毎のビット配分・空間解像度配分においては、大いに改善の余地が残されている。また H.264 符号化においては二本の 4:2:0 シーケンスを独立に符号化したため、動きベクトル情報やモード選択情報に冗長性がある。この考察も今後の課題である。

表 4: 多バンド映像符号化からみた各符号化方式の比較

符号化方式	符号化バンド数	色差フォーマット	ピクチャタイプ
MPEG-4 Studio Profile	3~6	4:2:0 / 4:2:2 / 4:4:4(ベース) + 4 / 4:4 / 4:4:4(aux.)	I, P
JPEG 2000	1~16,384	任意	Iのみ
H.264	3のみ	4:2:0のみ	I, P, B (本実験では I, P)

表 5: MPEG-4 における符号化条件

Profile	Studio Profile
GOV 構造	IPPP..
MV 探索範囲	±8
Alternate scan	なし
量子化マトリクス	デフォルト
色差フォーマット	4:4:4 (+aux. 4:4:4)
aux_comp_type	全て 'user-defined'
量子化パラメータ	固定
量子化スケール	非線形

表 6: JPEG 2000 における符号化条件

マルチコンポーネント変換	不使用
可逆モード	不使用
JJ2000 COM マーカ	不使用
レイヤ	単一

表 7: H.264 における符号化条件

GOP 構造	IPPP..
MV 探索範囲	±8
CABAC	使用
参照フレーム数	5
量子化パラメータ	固定
R-D 最適化	使用

謝辞 本研究を進めるにあたり多バンド映像データを提供いただいた、通信・放送機構ナチュラビジョンリサーチセンタの藤川研究員、内山研究員に感謝します。

## 参考文献

- [1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Proc. Magazine*, vol. 19, no. 1, pp. 17-28, Jan. 2002.
- [2] CIE Div. 8 TC-07, "Multispectral imaging," Nov.

2002.

- [3] 羽石秀昭, 山口雅浩, and 大山永昭, "マルチスペクトルに基づく映像技術," in *Proc. PCSJ2003*, Nov. 2003, pp. 2-6.
- [4] ISO/IEC SC 29/WG 11 (MPEG), *Information technology - Coding of audio-visual objects - Part 2: Visual - Amendment 1: Studio profile*, vol. ISO/IEC 14496-2:2001/Amd.1:2002, International Standard, Feb. 2002.
- [5] ISO/IEC SC 29/WG 1 (JPEG), *JPEG 2000 image coding system*, vol. ISO/IEC 15444-1:2000, International Standard, Dec. 2000.
- [6] ISO/IEC SC 29/WG 11 (MPEG), *Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding*, vol. ISO/IEC 14496-10:2003, International Standard, 2003.
- [7] Th. Keusen, "Multispectral color system with an encoding format compatible with the conventional tristimulus model," *Journal of IS&T*, vol. 40, no. 6, pp. 510-515, Nov./Dec. 1996.
- [8] 礮/富士写真フイルム, "マルチスペクトル画像の画像圧縮方法および画像圧縮装置," Apr. 2001, 特開 2001-119587.
- [9] S. Takamura and Y. Yashima, "Basic investigation of multiband video coding," in *Proc. PCS2003*, Apr. 2003, pp. 539-542.
- [10] EPFL, Canon Research France, and Ericsson, "JJ2000 Java implementation of JPEG 2000 ver. 5.1," Aug. 2002, <http://jj2000.epfl.ch/>.
- [11] "H.264 reference software version JM 7.3," Aug. 2003, <http://bs.hhi.de/~suehring/tml/>.