

# [特別招待講演] 高臨場感通信のための次世代映像符号化

坂東 幸浩<sup>1</sup>

**概要:** 本研究は、実物と見分けの付かない究極の高画質映像（ナチュラルクオリティ映像と呼ぶ）を用いた高臨場感通信に関する検討の一環として行われたものである。ナチュラルクオリティ映像を実現するためには、時間解像度、空間解像度、ビット深度、色再現性の観点から映像の表現能力を高める必要がある。しかし、ナチュラルクオリティ映像に要求される高い表現能力は、データ量の爆発的な増大を招き、その伝送・蓄積を困難にしていた。そこで、筆者らは、ナチュラルクオリティ映像の高画質を保持しつつ、そのデータ量を圧縮する効率的な符号化技術について検討してきた。本稿では、ナチュラルクオリティ映像に対する符号化技術として、時間解像度を高めた「高フレームレート映像」、ビット深度を高めた「高ビット深度映像」、空間解像度を高めた「高精細映像」、各々に対する検討結果を概観する。

## 1. はじめに

近年、ネットワーク環境の充実に伴い、高い臨場感・質感を備えた高画質映像による高臨場感映像通信への期待が高まっている。高画質映像の実現には、即ち、空間解像度、画素値深度、色再現性、時間解像度の向上が必要である。この4つの観点から、高画質化をはかり、実物と区別のない臨場感・質感を備えた映像をナチュラルクオリティ映像と呼ぶことにする。近年の半導体技術の進展を受け、ナチュラルクオリティ映像を実現する環境が整いつつある。

一方、ナチュラルクオリティ映像を用いて、高臨場感通信を実現する際の課題は、同映像を表現するために必要となる膨大なデータ量への対応である。こうしたデータ量の増大は、伝送路帯域・蓄積容量の逼迫を引き起こすためである。このため、増加したデータ量を圧縮によりコンパクトに表現し、ナチュラルクオリティ映像を効率的に通信するための符号化方式が必要となる。

本稿では、前述の空間解像度、時間解像度、画素値深度の観点から、画質を高めた映像信号を対象として、高画質化実現のための関連技術動向を整理すると共に、筆者の研究グループで取組んできた高画質映像に対する符号化研究について紹介する。まず、**2.**において、高フレームレート化に特有のデータ増大への対応として、その情報理論的性質の精査に基づき、フレームレートとフレーム間予測誤差の符号量の関係を定量的に示した理論モデルを概説する。次に、**3.**において、高ビット深度映像に対する符号化器設計法を紹介する。本手法の特徴は、ビット深度変換処理を

ビット深度方向の予測処理とみなし、その予測誤差の最小化に基づき、ビット深度変換処理を設計した点にある。さらに、**4.**において、高精細映像符号化において重要となる直交変換の次元数に着目し、直交変換の次元数と符号化効率の関係を定量的に示した理論モデルを概説する。あわせて、4Kサイズの映像をリアルタイムで符号化処理可能なH.264/AVC スケーラブル拡張規格 SVC 準拠のソフトウェアコーデックを紹介する。デジタルシネマ（4K 映像）や Super Hi-Vision（8K 映像）に代表される高精細映像の広がりと共に、筆者らは、高精細映像向けスケーラブル符号化の重要性を認識し、上記ソフトウェアコーデックを開発した。

## 2. 高フレームレート映像符号化に関する研究

時間解像度の向上、即ち、映像の高フレームレート化に関しては、撮像系の開発が先行する形で、検討が行われてきた。これは、計測応用としての高速度カメラ開発が進められてきたためである。応用例としては、FA 装置の不良動作解析、衝撃・落下試験、スポーツ選手のフォーム分析等がある。高速度カメラの持つ高い時間分解能を利用した物体追跡等の画像処理応用 [1] [2] も検討されている。また、高速度カメラに関する興味深い開発事例としては、高精細化と高フレームレート化を同時に実現した例として、ハイビジョンの順次走査で毎秒 300 フレームを撮影可能な高速度カメラ [3]、撮影時の工夫で高速撮影を実現した例として、低スペックなカメラ (30hz) から成るカメラアレイを構成することで 300[Hz] の高速度撮影を実現した撮像システム [4] 等がある。ただし、こうした高速度カメラは、撮影し

<sup>1</sup> NTT アドバンステクノロジー株式会社  
川崎市川崎区駅前本町 12-1

た映像のスロー再生を前提とした撮影機器であり、実時間表示における動画質向上を目的としたものではなかった。

本研究の開始当初(2005年)、映像の高画質化はポストHDTVとして4K映像を目指す空間解像度の向上に主眼がおかれており、フレームレートについては50~60Hzで十分という暗黙の認識があった。液晶ディスプレイのホールド効果に起因する動画質低下を改善する方策として、フレームレート向上の必要性は提起されて始めていた。ただし、あくまでも、ホールド効果という液晶ディスプレイ特有の問題に対する回避策としての検討であった。これに対し、筆者は、こうしたフレームレート(50-60Hz)がフリッカー(画面の明滅)の検知限界に基づき設定された点に着目し、滑らかな動きの表現には、より高いフレームレートを必要とするという仮説から研究をスタートした。まず、この仮説を検証すべく、高フレームレート映像の実視聴環境の構築を行い、高フレームレート化による動きの絶対的滑らかさを実証し、高フレームレート化が高画質化に不可欠な要素であることを証明した。その後、高速度カメラとともに、240[Hz]で描画可能なLCD、300[Hz]で描画可能なFEDも開発されており、こうしたディスプレイを用いて、フレームレートの増加に伴う画質向上が検証されている[5][6]。

一方、こうした映像の高フレームレート化はデータ量の増加を招く。そこで、筆者らは、フレームレートの増加に伴う発生符号量の変動を定量的に評価する目的から、発生符号量に影響を及ぼすフレーム間予測誤差信号に着目し、フレーム間予測誤差信号とフレームレートの数理的な関係を示す理論モデルを解析的に導出した<sup>7</sup>。同理論モデルにより、「フレームレートの増加に伴い、フレーム間予測誤差信号の符号量がどのように変動するのか?」という問いに対して、定量的な解を示すことが可能となった。

映像信号の時間軸方向の特徴を表す要素として、フレームレートのほかに撮像系の開口時間がある。開口時間の増加は、フレームスキップに起因するジャーキネスの低減に有効である。一方、開口時間の増加は、撮像デバイス蓄積時の積分効果により動きボケを引き起こす。このように、開口時間により、ジャーキネスと動きボケのトレードオフを制御可能である。また、開口時間の増加に伴う積分効果は低域通過特性を示す。こうした撮像系の低域通過特性は映像信号の性質を変化させる。このため、映像信号に対するフレームレートの影響を考察する際、開口率は重要な要素となる。

そこで、筆者らは、上記モデルを一般化し、撮像系の開口時間に応じて発生する低域通過特性の影響を考慮した形で、フレームレートとフレーム間予測誤差信号の情報量の関係を表す理論モデルを解析的に導出した[7]。実画像を用いた符号化実験の結果、同モデルの理論値が実測値に対する良好な近似値を与えることを確認し、同モデルの妥当

性を示した。また、同モデルが動画像の物理的な特性と適合していることも合わせて示した。上述の一連の研究の貢献は、従来、定性的な理解に留まっていた符号化特性に対して、数理的な根拠を与えた点にある。

### 3. 高ビット深度映像符号化に関する研究

忠実な画像表現の観点から、高ビット深度画像が注目を集めている。まず、白飛び・黒つぶれといった画質劣化を防ぐには、表示デバイスのダイナミックレンジを拡張する必要がある。さらに、そうしたダイナミックレンジの拡大後に、擬似輪郭の発生を回避するためには、十分な階調数を表現可能なビット深度が必要となる。なお、人の視覚系の検知限界は、12~13 [bits/channel] [8] [9] とされている。これが、映像の高ビット深度化が求められる所以である。なお、IEEE Signal Processing Society のフラグシップコンファレンス ICIP2010 にて高ビット深度・高ダイナミックレンジ画像処理に関する Special session を企画されており、同企画のサーベイ論文 [10] により、最新の技術動向を掴むことができる。

映像の高ビット深度化に関しては、カメラ・ディスプレイの高ビット深度化が進み、さらに、そうした高ビット深度対応のカメラとディスプレイ間の接続インターフェースとして、16[bits/channel] に対応した Displayport, HDMI 1.3 が規格化されたことを受け、高ビット深度映像を表現する土台は整いつつある。こうした撮影・表示技術とともに、効率的な伝送・蓄積のための符号化技術が必要となる。

画像・映像符号化の標準規格においても高ビット深度対応が進められている。色サンプルフォーマット 4:4:4 を対象として、AVC/H.264 は3つの Profile(High 4:4:4 predictive, High 4:4:4 intra, CAVLC 4:4:4intra)において最大14[bits/channel]をサポートしている。また、JPEG2000 は最大12[bits/channel]、JPEG-XR は最大32[bits/channel]を、各々、サポートしている。しかし、符号化ツールに関しては、既存の8[bits/channel]の信号に対する符号化ツールの転用であり、高ビット深度映像の特性を考慮した設計はなされていない。

高ビット深度映像処理の要素技術に tone mapping と呼ばれる階調変換技術がある。これは、高ビット深度映像(例:12 [bits/channel])を現行ディスプレイ(ビット深度8[bits/channel])に表示するために映像のビット深度を変換する技術である。高ビット深度映像に特有のビット深度方向の冗長性を除去する方法として、この tone mapping の符号化への応用が検討されている [11] [12] [13] [14] [15]。しかし、そこで用いる tone mapping はディスプレイ表示のための階調変換技術の転用であったため、本質的なビット深度方向の冗長性除去には至っていなかった。

これに対し、筆者らは、tone mapping をビット深度方向の予測処理と捉え、予測誤差最小化の観点から tone map-

ping を設計した。同設計における特徴は、適応量子化手法を利用し、ビット深度方向の予測機構を最適化した点にある [16]。さらに、拡張として、同予測機構の最適化を領域分割ベースの空間的な適応処理を導入することで、予測誤差の低減を図る手法を検討した。このとき、同適応処理の導入は、ビット深度変換処理の付加情報の増加をもたらすため、同時に、同付加情報に対する効率的な表現方法を導入した。[17]。この結果、高ビット深度映像に対して、H.264/AVC 比で符号化効率を 10%以上向上させることに成功した。

#### 4. 高精細映像符号化に関する研究

ディスプレイの大画面化に伴い、HDTV の次を見据えた映像の高精細化に関する検討が進められている。代表的な例としては、4K(4096 × 2160 [画素/フレーム]) 映像であるデジタルシネマ [18]、8K(7680 × 4320 [画素/フレーム]) 映像である SHV [19] があげられる。

高精細映像に対する符号化効率向上のアプローチとしては、符号化処理ブロックサイズおよび直交変換ブロックサイズの拡大があげられる。前者に関しては、8K,4K 映像に対する AVC/H.264 の規格拡張として、マクロブロックサイズの拡張 [20]、[21] が検討されている。後者に関しては、AVC/H.264 FRExt [22] における検討において、HDTV サイズの映像信号に対しては、4 × 4 次元の DCT に加えて、8 × 8 次元の DCT を用いる事で、符号化効率が向上する事が報告されている。さらに、HDTV を超える解像度の映像信号に対しては、4 × 4 次元および 8 × 8 次元の DCT とあわせて、16 × 16 次元の DCT を併用することで符号化効率が向上することが報告されている [23] [24]。このように、画像の精細化に伴い、変換符号化の次元数を増加させることで、符号化効率が向上することは経験的には知られていた。しかし、変換符号化の次元数と符号化効率の関係は、定性的な理解に留まり、理論的な根拠を与えるための解析は十分なされていない状況にあった。

そこで、筆者らは、KL 変換の次元数と情報集約性に対する数理モデルを構築し、両者の定量的な関係 [25] を明らかにした。なお、ここでは、KL 変換が変換後の情報の偏在度に関して最適な変換基底を与える変換であることから、同変換を理論的な解析対象とした。KL 変換について次元数と変換効率の関係をモデル化する際の難しさは、KL 変換後の状態では、観測可能な情報は固有値に限定され、評価したい情報間の関係が直接的には観測できない点にある。つまり、KL 変換後の情報の偏在度は、固有値という形に集約される。従って、そうした偏在度と KL 変換の次元数との関係を議論するためには、固有値という形に凝縮された情報に対して、その内部に潜む KL 変換の次元数と入力信号の関係を観測可能な形で定式化する必要がある。そこで、この問題を解決するための数学的道具として、量

子情報理論における密度作用素と von Neumann エントロピーを用いた。KL 変換の分散共分散行列を形式的に密度作用素とみなし、von Neumann エントロピーの性質を利用することにより、次元数の異なる KL 変換後の情報の偏在度を定量的に評価することが可能となった。

オリジナル映像はナチュラルクオリティ映像として保存し、映像の配信先で視聴者の表示端末・ネットワーク環境に応じて限定的なスペックで映像を表示する場合、単一の符号化データから所望のスペックの映像信号を復号可能なスケーラブル符号化が有効となる。4K 映像・8K 映像等の高精細映像の普及が進展したことを受け、こうした高精細映像を対象とした空間解像度スケーラビリティ機能の重要性が高まっていた。しかし、スケーラビリティ機能の付与は符号化効率の低下・処理量の増加を招く。H.264/AVC のスケーラブル拡張規格 SVC は、元々高負荷な H.264/AVC に、さらにスケーラビリティ機能を付与したため、HD サイズ以上の映像に対して、ソフトウェアによるリアルタイム符号化処理は実現困難とされていた。これに対し、筆者の研究グループは、SVC における階層的な処理構造を利用することで、符号化効率向上のみならず、符号化処理の低演算量化も同時に実現できることを見出し [26]、フル HD 対応 SVC リアルタイムソフトウェアコーデックの開発に成功した。同コーデックは、世界最大の放送機器展 IBC2008 に出展された。さらに、その拡張として、4K 映像対応 SVC リアルタイムソフトウェアコーデック [27] の開発にも成功し、同コーデックは、世界最大の映像機器展 NAB2009 に出展された。また、同コーデックは、2009 年 7 月の皆既日食のライブ伝送実験 [28] において、日食の瞬間を高臨場に体感するためのコア技術として使用された。

#### 5. おわりに

今後、コミュニケーションにおける映像の重要性は、益々高まると見込まれる。その際、必要な映像の高画質化を支える技術は、撮影・表示・伝送と多岐に渡る。こうした広範囲の技術分野に跨る取組は、産業界を活性化させる可能性を秘めていると考える。高画質映像に関する符号化研究が、そうした活性化のきっかけの一つになれば、筆者にとっては望外の喜びである。微力ではあるが、今回の受賞を励みとして、世の中を元気づけるような研究成果の創出に向けて努力を重ねる所存である。

**謝辞** これまで研究を進めるにあたり、親身に御指導を頂いた先輩諸氏・共同研究者の皆様、また、熱心に議論下さった NTT メディアインテリジェンス研究所の皆様から感謝申し上げます。

#### 参考文献

- [1] U. Muehlmann, M. Ribo, P. Lang and A. Pinz: A new high speed CMOS camera for real-time tracking appli-

- cations, *IEEE Int. Conf. on Robotics and Automation*, pp. 5195–5200 (2004).
- [2] 小室 孝, 石井 抱, 石川 正俊, 吉田 淳: 高速対象追跡ビジョンチップ, *信学論*, Vol. J84-D-II, pp. 75–82 (2001).
- [3] 小笠原 俊英, 山内 正仁, 戸村 義男, 山崎 順一, 後藤 正勝, 橋本洋二, 長 秀雄, 高地 栄一, 金山茂弘: 毎秒300フレーム順次走査HDTV高速度カメラ, *映メ誌*, Vol. 60, No. 3, pp. 358–365 (2007).
- [4] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz and M. Levoy: High Performance Imaging Using Large Camera Arrays, *ACM Trans. on Graphics*, Vol. 24, No. 3, pp. 765–776 (2005).
- [5] Y. Kuroki, T. Nishi, S. Kobayashi, H. Oyaizu and S. Yoshimura: A psychophysical study of improvements in motion-image quality by using high frame rates, *Journal of the Society for Information Display*, Vol. 15, No. 1, pp. 61–68 (2007).
- [6] 大村耕平, 菅原正幸, 野尻裕司: 撮像のフレーム周波数・開口率をパラメータとしたジャーキネスの主観評価, *信学技法*, Vol. 108, No. 425, IE2008-205, pp. 7–11 (2009).
- [7] Y. Bandoh, S. Takamura, H. Jozawa and Y. Yashima: Generalized theoretical modeling of inter-frame prediction error for high frame-rate video signal considering integral phenomenon, *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E93-A, No. 8, pp. 1442–1452 (2010).
- [8] T. Fujine, T. kanda, Y. Yoshida, M. Sugino, M. Teragawa and N. Ohta: Bit Depth Needed for High Image Quality TV-Evaluation Using Color Distribution Index, *Journal of Display Technology*, Vol. 4, No. 3, pp. 340–347 (2008).
- [9] R. L. Heckaman and M. D. Fairchild: Beyond the Locus of Pure Spectral Color and the Promise of HDR Display Technology, *Information Display*, Vol. 23, No. 7, pp. 22–26 (2007).
- [10] Y. Bandoh, G. Qiu, M. Okuda, S. Daly, T. Aach and O. C. Au: Recent advances in high dynamic range imaging technology, *Proc. of IEEE Int. Conf. Image Processing*, pp. 3125–3128 (2010).
- [11] G. Ward and M. Simmons: JPEG-HDR: A backward-compatible, high dynamic range extension to JPEG, *IEEE Int. Conf. on Image Processing*, pp. 283–290 (2005).
- [12] R. Mantiuk, A. Efremov, K. Myszkowski and H. Seidel: Backward Compatible High Dynamic Range MPEG Video Compression, *ACM Trans. Graph.* (2006).
- [13] M. Chen, G. Qiu, Z. chen and C. Wang: JPEG Compatible Coding of High Dynamic Range Imagery using Tone Mapping Operators, *Proc. 25th Picture Coding Symposium 2006* (2006).
- [14] M. Winken, H. Schwarz, D. Marpe and T. Wiegand: SVC bit depth scalability, *JVT-V078*, Marrakech, Morocco (2007).
- [15] M. Okuda and N. Adami: Two-Layer Coding Algorithm For High Dynamic Range Images based on Luminance Compensation, *Journal of Visual Communication and Image Representation*, Vol. 18, No. 5, pp. 377–386 (2007).
- [16] 伊藤健, 坂東幸浩, 高村誠之, 如澤裕尚, 上倉一人, 八島由幸: ビット深度変換処理を用いた高ビット深度画像の符号化に関する基礎検討, *信学論*, Vol. J93-D, No. 9, pp. 1658–1660 (2010).
- [17] 伊藤健, 坂東幸浩, 高村誠之, 如澤裕尚: 適応ビット深度変換処理を用いた高ビット深度映像符号化技術, *信学論*, Vol. J95-D, No. 11, pp. 1927–1934 (2012).
- [18] 中嶋正之他: 小特集 デジタルシネマ, *映メ誌*, Vol. 59, No. 2, pp. 199–217 (2005).
- [19] Nojiri, Y.: An approach to Ultra High-Definition TV, *International symposium on universal communication*, No. 5-1 (2007).
- [20] S. Sakaïda, K. Iguchi, N. Nakajima, Y. Nishida, A. Ichigaya, E. Nakasu, M. Kurozumi and S. Gohshi: The Super Hi-Vision Codec, *IEEE Int. Conf. on Image Proc. 2007*, Vol. 1, pp. I-21–I-24 (2007).
- [21] 杉本 和夫, 関口 俊一, 山岸 秀一, 山田悦久: 高解像度映像信号向け動画像符号化方式に関する一検討, *PCSJ2009*, P-3-22 (2009).
- [22] G. Sullivan, P. Topiwala and A. Luthra: The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions, *Proc. SPIE Applications of Digital Image Processing XXVII*, Vol. 5558, pp. 454–474 (2004).
- [23] S. Sakaïda, A. Ichigaya, E. Nakasu and S. Gohshi: Compression Coding of Super Hi-Vision via Expansion of AVC/H.264 Transform Block Size, *International Workshop on Advanced Image Technology (IWAIT)*, P1-24, pp. 902–907 (2007).
- [24] T. Yamamoto and T. Ikai: A new intra macroblock type for adaptable transform block size, *ITU-T Q.6/SG16 Doc.*, Vol. COM 16-C 88-E (2009).
- [25] 大林浩気, 坂東幸浩, 高村誠之, 如澤裕尚, 八島由幸: Karhunen-Loeve 変換の情報集約性に関する理論的考察, *信学論*, Vol. J93-A, No. 9, pp. 626–637 (2010).
- [26] 早瀬和也, 坂東幸浩, 藤井寛, 高村誠之, 如澤裕尚, 上倉一人: H.264/AVC 拡張規格 SVC 準拠フル HDTV 映像伝送システム, *信学論*, Vol. J93-D, No. 10, pp. 2177–2187 (2010).
- [27] 早瀬和也, 坂東幸浩, 藤井寛, 高村誠之, 上倉一人, 如澤裕尚: 4K×2K 対応 SVC リアルタイムソフトウェアコーデック, *信学総大* (2010).
- [28] 尾久土正己: 4K 映像システムを使った皆既日食の全天投影, *映像情報メディア学会誌*, Vol. 63, No. 10, pp. 1385–1389 (2009).