

高フレームレート映像信号に対する符号化特性解析

坂東幸浩[†], 高村誠之[†], 上倉一人[†], 八島由幸[†]

[†] 日本電信電話株式会社 NTT サイバースペース研究所

〒 239-0847 神奈川県横須賀市光の丘 1-1

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

あらまし 近年、臨場感あふれる大画面のスポーツ映像やデジタルシネマに代表される超高画質映像への期待が高まっている。しかし、被写体の自然な動きを表現するために不可欠な時間解像度の向上、即ち、映像の高フレームレート化については、十分な検討がなされていない。現行映像のフレームレートはフリッカーの検知限から設定されており、自然な動き表現には、さらに高いフレームレートが必要となる。しかし、高フレームレート化により高画質化を追求しようとした場合、データ量の増大に対応するために効率的な符号化方式が必要となる。そこで、本報告では、高フレームレート映像の符号化を睨み、フレームレートと情報量の関係について定量的な評価を行うことを目的とし、フレームレートとフレーム間差分信号の情報量の関係を示す解析的なモデルを導出した。実画像を用いた符号化実験を通して、同モデルの理論値が、実画像の符号化結果に対する適切な近似値となっていることを確認した。

Encoding characteristic analysis of high frame-rate video signal

Yukihiro BANDO[†], Seishi TAKAMURA[†], Kazuto KAMIKURA[†], and
Yoshiyuki YASHIMA[†]

[†]NTT Cyber Space Laboratories, NTT Corporation

1-1 Hikarino-oka, Yokosuka, Kanagawa 239-0847, JAPAN
E-mail : {bandou.yukihiro, takamura.seishi, kamikura.kazuto,
yashima.yoshiyuki}@lab.ntt.co.jp

Abstract:

Realistic representations using extremely high quality images are becoming increasingly popular. For example, digital cinemas can now display moving pictures composed of high-resolution digital images. Although these applications focus on increasing the spatial resolution of the images, they ignore the temporal resolution. In order to faithfully represent the movement of an object in a video sequence, we need high-frame-rate images whose statistical properties have to be elaborated. This paper establishes for high frame-rate video a mathematical model of the relationship between frame-rate and bit-rate. A coding experiment confirms the validity of the mathematical model.

1 はじめに

近年、臨場感あふれる大画面のスポーツ映像やデジタルシネマ [1] に代表される超高画質映像への期待が高まっている。これを受けて、映像の高精細化に関する研究が精力的に行われている [2]。超高画質映像の実現には次の四要素が必要である。すなわち、空間解像度、画素値深度、色再現性、時間解像度である。これを受けて、前者三要素についてはデジタルシネマ等の応用およびナチュラルビジョンプロジェクト [3] において検討が進められている。

しかし、被写体の自然な動きを表現するために不可欠な時間解像度の向上、即ち、映像の高フレームレート化については、十分な検討がなされていない。Spillmann らによるによれば、網膜の出力細胞である神経節細胞が出力するパルス数の上限は毎秒 300 ~ 400 個程度であるとする生理学的な知見が示されている [4]。このため、人の視覚系は 1/150 ~ 1/200 秒程度までの発光の違いを知覚できると推察される。これは、知覚可能なフレームレートの検知限が、150 ~ 200 [フレーム/秒] であることを意味する。現行映像のフレームレートである 30,60 [フレーム/秒] は、フリッカーの検知限から定められたものであり、自然な動きを表現するに十分な値ではない。

映像の超高画質化はデータ量の爆発的な増加を招くため、効率的な符号化法が不可欠となる。高フレームレート映像の符号化を検討する際、同映像信号の統計的性質を的確に把握することが必要となる。とりわけ、フレームレートと情報量の関係は重要である。フレームレートの増加に伴い、映像シーケンス全体でみた場合の情報量は増大する。一方、フレーム間の相関は大きくなるため、フレームあたりの情報量で見れば、減少することが予想される。しかし、その定量的な性質は把握されていない。また、従来研究として、フレームレートと情報量の関係に注目した可変フレームレートによる時間スケラビリティの検討 [5, 6, 7] が行われているが、60 [フレーム/秒] 以下の低フレームレート映像に関する内容が中心であり、高フレームレート映像信号の統計的性質についての知見は得られていない。

そこで、本報告では、高フレームレート映像信号の符号化を睨み、フレームレートとフレーム間予測誤差信号の情報量を結び付ける解析的なモデルの導出を行う。また、実画像を用いた符号化実験を通して、同モデルの妥当性を検証する。

2 フレームレートと情報量の関係

2.1 解析的なモデルの導出

フレームレートと情報量の関係を解析的に導出する。ここでは、簡単のため 1 次元信号を例にとり説明する。時刻 t の位置 x における信号 $f_t(x)$ に対して、区間 B を単位として動き補償 (推定変位量 \hat{d}) を行った場合、その区間内の動き補償後の予測誤差は次のように表現できる。

$$\begin{aligned}\sigma^2 &= \sum_{x \in B} |f_t(x) - f_{t-1}(x + \hat{d})|^2 \\ &= \sum_{x \in B} |f_{t-1}(x + d) - f_{t-1}(x + \hat{d}) + n|^2 \\ &= \sum_{x \in B} \left| \left\{ \frac{d}{dx} f_{t-1}(x) \right\} (d - \hat{d}) + \phi(x) + n \right|^2 \\ &\simeq a(d - \hat{d})^2 + b(d - \hat{d}) + N\end{aligned}\quad (1)$$

ここで、 a, b, N は以下の通りである。

$$\begin{aligned}a &= \sum_{x \in B} \left\{ \frac{d}{dx} f_{t-1}(x) \right\}^2 \\ b &= 2 \sum_{x \in B} \left\{ \frac{d}{dx} f_{t-1}(x) \right\} \phi(x) \\ N &= \phi(x)^2 + n^2\end{aligned}$$

また、 d は真の変位量、 $\phi(x)$ はテイラー展開の 2 次以降の項、 n および N は雑音成分に起因する項とした。式 (1) への近似では、雑音成分の独立性、およびテイラー展開の一次近似を利用した。さらに、フレーム間での動きを等速運動とみなせば、変位量はフレーム間隔 (フレームレートの逆数) に比例することから、フレームレート F とは反比例の関係にある。このため、 $d - \hat{d}$ は次式のように表せる。

$$d - \hat{d} = \kappa F^{-1}$$

ここで、 κ は画像に依存した定数である。上式を式 (1) に代入し、次式を得る。

$$\sigma^2 = AF^{-2} + BF^{-1} + N\quad (2)$$

ここで、 A, B は以下のようにおいた。

$$\begin{aligned}A &= \kappa^2 a \\ B &= \kappa b\end{aligned}$$

予測誤差がラプラス分布に従うと仮定し、同分布の differential entropy を用いて、予測誤差の情報量

$I(F)$ を見積もると次式の関係を得る。

$$\begin{aligned} I(F) &= \log(\sigma) \\ &= \frac{1}{2} \log(AF^{-2} + BF^{-1} + N) \quad (3) \end{aligned}$$

2.2 モデルの特性

式 (3) のモデルは次式の関係満足する。

$$\lim_{F \rightarrow \infty} = \frac{1}{2} \log(N) \quad (4)$$

$$\frac{dI(F)}{dF} < 0 \quad (5)$$

式 (4) は、フレーム間隔が極限まで小さくなる（フレームレートが極限まで大きくなる）と、フレーム間予測誤差信号の情報量として雑音成分の情報量が支配的になることを示しており、映像信号の物理的な性質に符合する。式 (5) は

$$\frac{dI(F)}{dT} > 0$$

と等価である。ここで、 T は隣接フレーム間の時間間隔であり、フレームレート F の逆数で表わされる。つまり、フレーム間隔の増大（フレームレートの減少）と共にフレーム間差分信号の情報量も増加することを示している。このように、式 (3) のモデルは、増加率のみならず、極限值に関しても動画像の物理的な特性と適合していることが確認できる。

3 実験

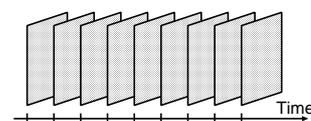
3.1 高フレームレート映像ソース

高フレームレートの映像ソースは、表 1 に示す NAC 社製ハイスピードカメラ（MEMRECOM fx RX-3）を用いて、デジタルデータとして作成した。撮影時のフレームレートは 1000 [フレーム/秒] した。また、シャッター時間はフレーム間隔に等しいものとした。さらに、デジタルデータとして出力する際、画素値にはガンマ補正・ニー補正といった非線形変換は施していない。

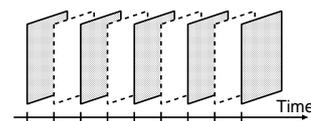
フレームレートと情報量の対応関係を求めるためには、画角が等しくフレームレートのみ異なる映像が必要となる。ここでは、図 1 に示すように、フレーム間引きにより、フレームレートの異なる映像を得る。同図 (b) は基準となる同図 (a) のシーケンスが

表 1: ハイスピードカメラ (NAC 社製 MEMRECOM fx RX-3) のスペック

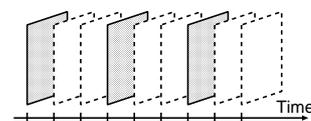
センサー素子	カラー CMOS
素子画素数	1280 × 1024 [画素]
画素値深度	10 [ビット/画素]
フレームレート	最大 1000 [フレーム/秒]
シャッター速度	1/1000 ~ 1/200000 [秒]
感度	ASA250 相当
レンズマウント	F マウント、C マウント



(a) 原画像シーケンス



(b) フレームレートを 1/2 に間引いたシーケンス



(c) フレームレートを 1/3 に間引いたシーケンス

図 1: フレームレートの変換

ら、1枚おきにフレームを間引き、フレームレートを半分にしたシーケンスを示しており、同図 (c) はフレームレートを 1/3 にしたシーケンスを示している。同図では、間引いたフレームを破線の矩形で示している。

フレーム間引きの基準となる映像は、表 1 の高速度カメラで撮影した RGB カラー映像であり、フレームレートは 1000 [フレーム/秒]、総フレーム数は 480 [フレーム]、解像度は 640 × 480 [画素] である。また、映像素材はテニスと野球のスイングを撮影したスポーツシーンである。各シーケンスは、時間方向の相関の強さが異なるものを選んでいく。

3.2 モデルの妥当性の検証

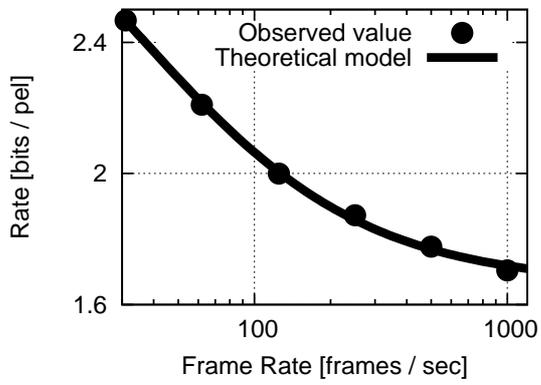
上述のモデルの妥当性を検証するために、実画像を用いた符号化実験を行った。図2にフレームレートを変化させた場合の情報量の変化を示す。破線は実画像による符号化結果を示しており、実線は式(3)のモデルによる理論値である。ここで、モデルのパラメータ (A, B, N) は最小二乗推定により求めた値を用いた。横軸はフレームレートを対数スケールで表わしている。縦軸は、映像の情報量として、動き補償誤差信号及び動ベクトルのエントロピの和を画素平均で表している。なお、動き補償時のブロックサイズは 16×16 [画素] とした。動き推定の精度は $\frac{1}{4}$ [画素] であり、小数画素位置の補間には MPEG-4 の補間フィルタ [8] を用いた。フレーム間予測は片方向予測とし、その際の参照フレームは直前のフレームとした。図2に示すように、導出したモデルの理論値はフレームレートの増加に伴うフレーム間予測誤差の情報量の減少を適切に表現できていることが確認できる。また、こうした理論値と実測値の関係は複数のコンテンツおよびシーンに対する符号化結果に対して確認できた。本実験結果は、モデルを導出する際の仮定、および、導出されたモデルの妥当性を裏付けるものと言える。

4 おわりに

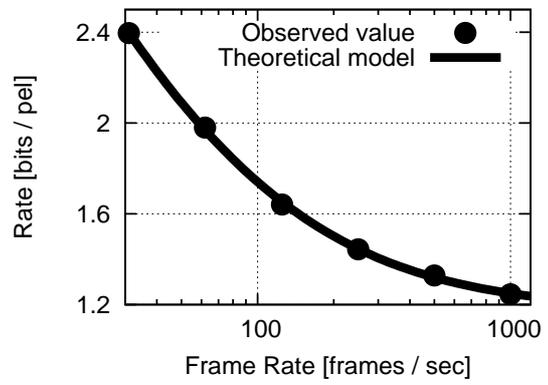
本報告では、フレームレートと情報量の定量的な関係を調査する目的から、両者の関係を表すモデルを解析的に導出した。実画像を用いた符号化実験の結果、同モデルの理論値が実測値に対する良好な近似値を与えることを確認し、同モデルの妥当性を示した。また、同モデルが動画の物理的な特性と適合していることも合わせて示した。本報告におけるフレームレートと情報量の間を関係を表すモデルは、将来の高フレームレート映像システムの構築に有益な指針を与えるものと考えられる。

参考文献

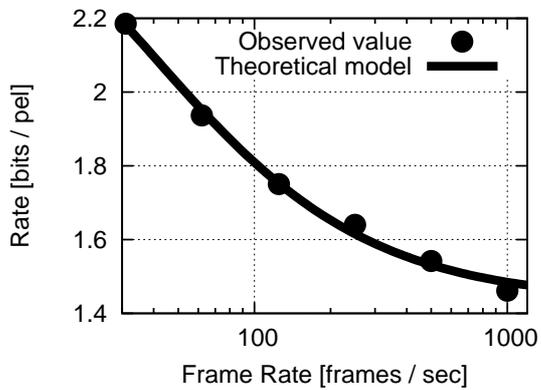
- [1] 中嶋正之他. 小特集 デジタルシネマの最新事情. 映メ誌, Vol. 57, No. 2, pp. 178–202, Feb. 2003.
- [2] 高畑文雄他. 小特集 大画面ディスプレイ. 映メ誌, Vol. 56, No. 8, pp. 1209–1232, Aug. 2002.
- [3] 羽石秀昭. マルチスペクトルに基づく映像技術. In *PCSJ2003*, pp. 1–6, 2003.
- [4] L.Spillmann and J.S.Werner. *Visual perception the neurophysiological foundations*. Academic Press, 1990.
- [5] 武川直樹, 川西久, 松岡毅. フレーム間符号化における駒落とし方式の検討. 信学論, Vol. J65-B, No. 4, pp. 455–462, Apr. 1982.
- [6] H. Song and C.-C.J.Kou. Rate control for low-bit-rate video via variable encoding frame rates. *IEEE trans. CSVT*, Vol. 11, No. 4, pp. 512–521, 2001.
- [7] 稲積泰宏, 吉田俊之, 酒井善則, 堀田裕弘. ビットレート制限下における動画通信のための最適フレームレートの推定. 信学論, Vol. J85-B, No. 7, pp. 1130–1142, 2002.
- [8] ISO/IEC JTC1/SC29/WG11 N4668. *Overview of the MPEG-4 Version 2 standard*, 2002.



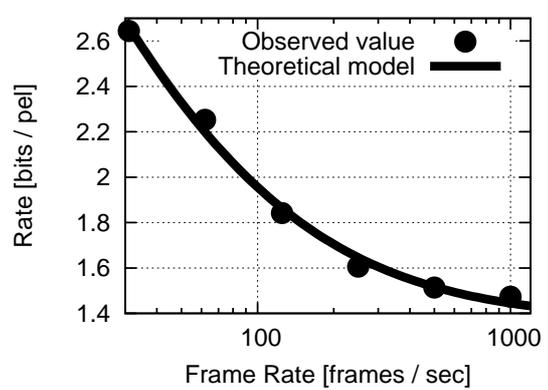
(a) Tennis [Scene 1]



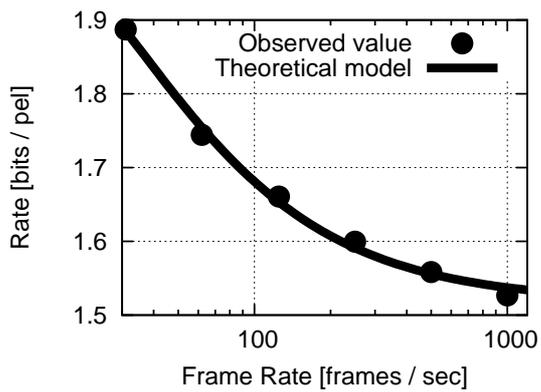
(b) Tennis [Scene 2]



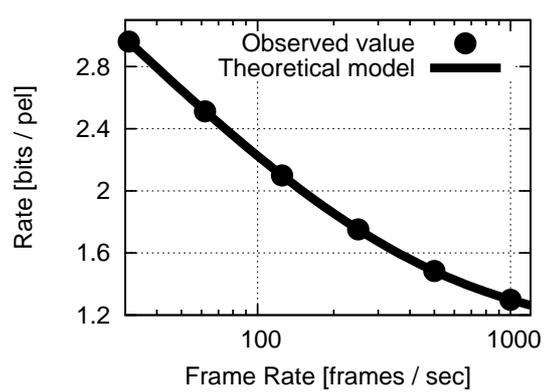
(c) Baseball [Scene 1]



(d) Baseball [Scene 2]



(e) Golf [Scene 1]



(f) Golf [Scene 2]

図 2: フレームレートと符号量の関係