

時間解像度変換時の低域通過特性を考慮した
 フレームレート・符号量モデルの構築
 Construction of frame-rate/bit-rate model combining temporal
 low-pass characteristic in frame-rate conversion

坂東幸浩[†], 上倉一人[†], 八島由幸[†]

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

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 [1]。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 らの示した生理学的な知見によれば、人の視覚系の知覚可能なフレームレートの検知限は 150 ~ 200 [フレーム/秒] であるとされる [2]。一方、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。

映像の高フレームレート化はデータ量の増加を招くため、効率的な符号化法が不可欠となる。その際、高フレームレート映像信号のフレーム間予測誤差信号の統計的性質を的確に把握することが必要となる。しかし、従来のフレーム間予測誤差信号のモデル化に関する検討 [3] では、フレームレートの影響が考慮されていなかった。こうした状況を鑑み、筆者らは、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について定量的な評価を行うことを目的として、両者の関係を示す理論モデルについて検討を行ってきた [4]。ただし、同モデルにおいては、フレームレートのダウンサンプリングが単純なフレーム間引きを想定したものであった。

そこで、本報告では、撮像系の開口時間に応じて発生する積分効果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する。あわせて、同モデルの妥当性を実画像を用いた符号化実験を通して検証する。

2 フレームレートと情報量の関係

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。位置 x において、シャッターを時刻 t から $t + \delta$ の間、開口して撮影した信号を $f_t(x, \delta)$ とする。このとき、シャッターの開口時間を $m\delta$ (m は自然数) に伸張した場合に得られる信号 $\bar{f}_{mt}(x, m\delta)$ は、 $f_t(x, \delta)$ を用いて以下のように表せる。

$$\bar{f}_{mt}(x, m\delta) = \frac{1}{m} \sum_{\tau=mt}^{m(t+1)-1} f_{\tau}(x, \delta) \quad (1)$$

つまり、フレームレートの変換に、シャッターの開口時間の変化が伴う場合は、式 (1) に示す低域通過フィルタの影響を考慮する必要がある。

信号 $\bar{f}_{mt}(x, m\delta)$ に対して、区間 L を単位として動き補償 (推定変位量 \hat{d}_m) を行った場合、同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。なお、以下では、位置 x における真の変位量を $d_m(x)$ としている。

$$\begin{aligned} \sigma_e^2 &= \sum_x |\bar{f}_{mt}(x, m\delta) - \bar{f}_{m(t-1)}(x + \hat{d}_m, m\delta)|^2 \\ &= \sum_x |\bar{f}_{m(t-1)}(x + d_m(x), m\delta) \\ &\quad - \bar{f}_{m(t-1)}(x + \hat{d}_m, m\delta) + \bar{n}(x, m\delta)|^2 \\ &= \sum_x \left| \frac{1}{m} \sum_{i=0}^{m-1} \{f_{m(t-1)+i}(x + d_m(x), \delta) \right. \\ &\quad \left. - f_{m(t-1)+i}(x + \hat{d}_m, \delta)\} + \bar{n}(x, m\delta) \right|^2 \\ &= \sum_x \left| \frac{\left\{ \sum_{i=0}^{m-1} \frac{d}{dx} f_{m(t-1)+i}(x, \delta) \right\}}{m} (d_m(x) - \hat{d}_m) \right. \\ &\quad \left. + \phi(x) + \bar{n}(x, m\delta) \right|^2 \quad (2) \end{aligned}$$

ここで、 $\phi(x)$ はテイラー展開の 2 次以降の項である。また、 $\bar{n}(x, m\delta)$ は撮像デバイスの熱雑音等に起因す

[†]日本電信電話株式会社 NTT サイバースペース研究所

る雑音成分である。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、フレームレートとは反比例の関係にある。このため、信号 $\bar{f}_{mt}(x, m\delta)$ のフレームレートを $F = (m \cdot \delta)^{-1}$ とおくと、 $d_m(x) - \hat{d}_m$ は次式のように表せる。

$$\begin{aligned} d_m(x) - \hat{d}_m &= (v_m(x) - \hat{v}_m) \cdot m \cdot \delta \\ &= (v_m(x) - \hat{v}_m) \cdot F^{-1} \end{aligned}$$

ここで、 $v_m(x)$ は位置 x の画素毎に定まる比例定数であり、 \hat{v}_m は区間 L に対して定まる比例定数である。両比例定数の物理的な意味はいずれも速度となる。

上式を式 (2) に代入し、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、次式を得る。

$$\sigma_e^2 \simeq A(m)F^{-2} + B(m)F^{-1} + N(m) \quad (3)$$

ここで、 $A(m), B(m), N(m)$ は以下の通りである。なお、以下では簡単のため、 $f_t(x, \delta)$ を $f_t(x)$ と略記する。

$$\begin{aligned} A(m) &= \sum_x \left\{ \varepsilon(x) \frac{1}{m} \sum_{i=0}^{m-1} \frac{d}{dx} f_{m(t-1)+i}(x) \right\}^2 \\ B(m) &= 2 \sum_x \left\{ \varepsilon(x) \frac{1}{m} \sum_{i=0}^{m-1} \frac{d}{dx} f_{m(t-1)+i}(x) \right\} \phi(x) \\ N(m) &= \sum_x \{ \phi(x)^2 + \bar{n}(x, m\delta)^2 \} \end{aligned}$$

なお、ここでは $\varepsilon(x) = v_m(x) - \hat{v}_m$ とおいた。

続いて以下では、 $A(m), B(m), N(m)$ とフレームレートの関係を示す。まず、 $\mu_{mt}(x) = \frac{1}{m} \sum_{i=0}^{m-1} f_{m(t-1)+i}(x)$ とおき、 $A(m)$ を次のように展開する。

$$\begin{aligned} A(m) &= \sum_x \left[\varepsilon(x) \frac{d}{dx} \mu_{mt}(x) \right]^2 \\ &\simeq \sum_x [\varepsilon(x) \{ \mu_{mt}(x) - \mu_{mt}(x-1) \}]^2 \\ &= \sum_x \{ \varepsilon(x) \mu_{mt}(x) \}^2 + \sum_x \{ \varepsilon(x) \mu_{mt}(x-1) \}^2 \\ &\quad - 2 \sum_x \{ \varepsilon(x)^2 \mu_{mt}(x) \mu_{mt}(x-1) \} \\ &\simeq 2 \sum_x \{ \varepsilon(x) \mu_{mt}(x) \}^2 \\ &\quad - 2 \sum_x \{ \varepsilon(x)^2 \mu_{mt}(x) \mu_{mt}(x-1) \} \\ &\simeq \beta \frac{2\sigma_s^2(1-\rho)}{m^2} \left\{ m - \frac{1-\rho}{\rho} \sum_{i>j} \alpha_{i,j} \rho^{|\bar{d}_i - \bar{d}_j|} \right\} \end{aligned} \quad (4)$$

ここで、 $\beta = \sum_x \varepsilon(x)^2$ とする。上述の近似では、 $\varepsilon(x)$ と $\mu_{mt}(x)$ は統計的に独立と仮定した。また、以下の斉次モデルを用いた。

$$\sum_x \{ f_t(x) \}^2 = \sigma_s^2$$

$$\sum_x \{ f_t(x) f_t(x+k) \} = \sigma_s^2 \rho^k$$

さらに、以下の近似を用いた。

$$\begin{aligned} &\sum_x \{ f_t(x + d_i(x)) f_t(x + d_j(x)) \} \\ &\simeq \alpha_{i,j} \sum_x \{ f_t(x + \bar{d}_i) f_t(x + \bar{d}_j) \} \\ &= \alpha_{i,j} \sigma_s^2 \rho^{|\bar{d}_i - \bar{d}_j|} \end{aligned}$$

ここで、 \bar{d}_i および \bar{d}_j は各々、 $d_i(x)$ および $d_j(x)$ ($x \in L$) の平均値である。 $\alpha_{i,j}$ は平均変移量 (\bar{d}_i, \bar{d}_j) を用いた近似に対する補正パラメータである。 ρ は画像信号の自己相関である。この ρ に着目して、式 (4) に対する近似を行う。画像信号の自己相関である ρ は 1 以下であり、かつ、1 に近い値とみなすことができる。このため、次の不等式を満たす。

$$\frac{1-\rho}{\rho} \ll 1.$$

上不等式により、式 (4) は以下のように近似できる。

$$A(m) \simeq \beta \frac{2\sigma_s^2(1-\rho)}{m}$$

さらに、 m はダウンサンプルされたフレームレート F の最大フレームレート $F_0 (= 1/\delta)$ に対する比なので、上式は、次式のように表すことができる。

$$A(m) \simeq \beta \frac{2\sigma_s^2(1-\rho)}{F_0} F \quad (5)$$

同様に、 $B(m)$ を次式のように近似する。

$$\begin{aligned} B(m) &\simeq 2\gamma \phi(x) \sqrt{A(m)} \\ &= 2\gamma \phi(x) \sqrt{\beta \frac{2\sigma_s^2(1-\rho)}{F_0} F} \end{aligned} \quad (6)$$

ここで、 γ は 1 あるいは -1 とする。

$N(m)$ における $\bar{n}(x, m\delta)$ について考察する。雑音成分 $\bar{n}(x, m\delta)$ は画像信号と統計的に独立だと仮定しているため、式 (1) の平均化処理は、雑音成分を低減させる。即ち、フレームレート $F_0 (= 1/\delta)$ の (4) シーケンスに含まれる雑音成分の分散を n_0 とする

と、 $\bar{n}(x, m\delta)^2$ は次式のように表される（詳細は 3.3 にて後述）。

$$\bar{n}(x, m\delta)^2 = \sum_{x \in L} \frac{n_0^2}{m} = \|L\| \frac{n_0^2}{F_0} F \quad (7)$$

ここで、 $\|L\|$ はセグメント L 内の要素数である。

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 $I(F)$ [bits/pel] を見積もると次式の関係を得る。なお、以下では、対数の底は 2 とする。

$$I(F) = \log(\sigma_e) = \frac{1}{2} \log(\hat{A}F^{-1} + \hat{B}F^{-1/2} + \hat{C}F + \hat{D}) \quad (8)$$

ここで、 \hat{A} , \hat{B} , \hat{C} , および \hat{D} は次式となる。

$$\begin{aligned} \hat{A} &= \beta \frac{2\sigma_s^2(1-\rho)}{F_0} \\ \hat{B} &= 2\gamma\phi(x) \sqrt{\beta \frac{2\sigma_s^2(1-\rho)}{F_0}} \\ \hat{C} &= \|L\| \frac{n_0^2}{F_0} \\ \hat{D} &= \sum_x \{\phi(x)\}^2 \end{aligned}$$

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、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。

フレームレート変換は、前述の高速度カメラで撮影した映像（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]）を入力として、式 (1) に基づき行った。また、映像素材はテニス、野球およびゴルフ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 1 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。黒丸は実画像による符号化結果を示しており、実線は式 (8) 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 (\hat{A} , \hat{B} , \hat{C} , \hat{D}) 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 (IS14496-2) の補間フィルタを用いた。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 1 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。

3.3 考察

図 1 より、フレームレートの増加に伴い、大局的には情報量が減少することが確認できる。これは、フレームレートの増加に伴い、フレーム間隔が狭まり、フレーム間の相関が大きくなることに由来する。一方、図 1(c) における 1000 [fps] と 500 [fps] との比較のように、フレームレートを小さくすることで情報量が減少する場合がある。これは、シャッターの開口時間の増加に伴い、雑音成分が抑圧されたことに起因する。フレームレートが減少すれば、動き補償誤差信号は大きくなる。一方、フレームレートの減少時には、シャッターの開口時間が増加するため、雑音成分が抑圧される。後者による情報量の減少が前者による情報量の増加を上回る場合、図 1(c) のような現象となって観測される。こうした現象は単純なフレーム間引きによってフレームレートを変化させていた場合には、発生しない現象である。

なお、シャッターの開口時間の増加に伴う雑音成分の抑圧は、以下のように説明できる。ここでは、画像信号 $f_t(x, \delta)$ が次式のように画像成分 $s_t(x, \delta)$ と雑音成分 $n_t(x, \delta)$ から構成されるものと仮定する。

$$f_t(x, \delta) = s_t(x, \delta) + n_t(x, \delta)$$

このとき、式 (1) は以下のように展開できる。

$$\begin{aligned}\bar{f}_{mt}(x, m\delta) &= \frac{1}{m} \sum_{\tau=mt}^{m(t+1)-1} \{s_t(x, \delta) + n_t(x, \delta)\} \\ &= \frac{1}{m} \sum_{\tau=mt}^{m(t+1)-1} s_t(x, \delta) + \frac{1}{m} \sum_{\tau=mt}^{m(t+1)-1} n_t(x, \delta)\end{aligned}$$

なお、上式の第 2 項を略記したものが式 (2) の $\bar{n}(x, m\delta)$ である。

さらに、雑音成分は時間的・空間的に無相関だとすると、上式は以下のように近似できる。

$$\begin{aligned}& \sum_x |\bar{f}_{mt}(x, m\delta)|^2 \\ &= \sum_x \left| \frac{1}{m} \sum_{\tau=mt}^{m(t+1)-1} s_t(x, \delta) + \frac{1}{m} \sum_{\tau=mt}^{m(t+1)-1} n_t(x, \delta) \right|^2 \\ &\simeq \frac{1}{m^2} \sum_x \left| \sum_{\tau=mt}^{m(t+1)-1} s_t(x, \delta) \right|^2 + \frac{1}{m^2} \sum_x \sum_{\tau=mt}^{m(t+1)-1} n_t(x, \delta)^2\end{aligned}$$

このとき、各フレームの雑音電力 $\sum_x n_t(x, \delta)^2$ を一定値 ν^2 と近似すると、上式は次式ようになる。

$$\sum_x |\bar{f}_{mt}(x, m\delta)|^2 = \frac{1}{m^2} \sum_x \left| \sum_{\tau=mt}^{m(t+1)-1} s_t(x, \delta) \right|^2 + \frac{\nu^2}{m}$$

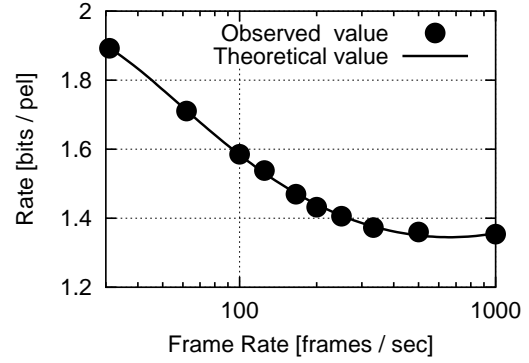
このように、雑音電力が $\frac{1}{m}$ に抑圧されることが分かる。なお、上式の ν^2 を画素あたりの平均値に直した値が式 (7) の n_0^2 である。

4 おわりに

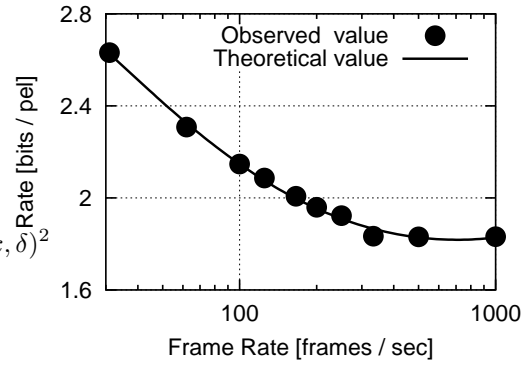
本報告では、フレームレートと情報量の定量的な関係を調査する目的から、両者の関係を表すモデルを解析的に導出した。実画像を用いた符号化実験の結果、同モデルの理論値が実測値に対する良好な近似値を与えることを確認し、同モデルの妥当性を示した。また、同モデルが動画像の物理的な特性と適合していることも合わせて示した。

参考文献

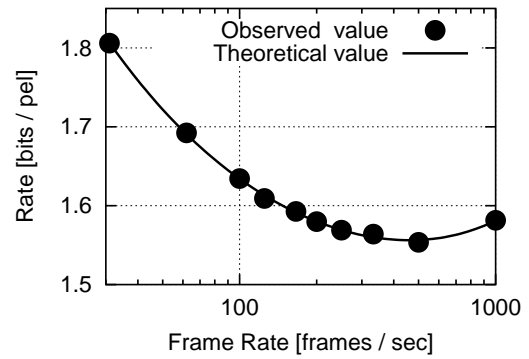
- [1] 中嶋正之他. 小特集 デジタルシネマ. 映メ誌, Vol. 59, No. 2, pp. 199–217, Feb. 2005.
- [2] L.Spillmann et.al., *Visual perception the neurophysiological foundations*. Academic Press, 1990.
- [3] Y.Shishikui. A study on modeling of the motion compensation prediction error signal. *IE-ICE Trans. Comm.*, E75-B, No. 5, pp. 368–376, 1992.
- [4] 坂東幸浩他. 高フレームレート映像信号の符号量に関する理論モデルの構築. *FIT2005*, J-074. 2005.



(a) Tennis



(b) Golf



(c) Baseball

図 1: フレームレートと情報量の関係