

開口率を考慮した高フレームレート信号に対する フレーム間予測誤差信号の理論解析 Theoretical analysis of inter-frame prediction error for high frame-rate video signal with variable aperture ratio

坂東幸浩[†], 早瀬和也[†], 高村誠之[†], 上倉一人[†], 八島由幸[†]
Yukihiro BANDO[†], Kazuya HAYASE[†], Seishi TAKAMURA[†], Kazuto
KAMIKURA[†], and Yoshiyuki[†] YASHIMA

1 はじめに

時間解像度は、空間解像度、画素値深度、色再現性と共に、映像の高画質化には不可欠な要素である。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。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 [1]。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り、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 [2] 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 [3]。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。

映像信号の時間軸方向の特徴を説明するために必要なパラメータとして、フレームレートのほかに撮像系の開口率がある。撮像デバイスへの入射光量を確保するには、一定以上の開口率が必要となる。開口率の増加は、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、開口率の増加は、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、開口率により、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、開口率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、本報告では、開口率に応じて発生する低域通過特性を考慮して、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。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。さらに、フレーム間予測誤差信号の情報量を最小化する開口率を導出する。

2 フレームレート-情報量モデル

フレームレートおよび開口率に対する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 $\bar{f}_t(x, m\delta)$ は、 $f_t(x, \delta)$ を用いて以下のように表せる。

$$\bar{f}_t(x, m\delta) = \frac{1}{m} \sum_{i=0}^{m-1} f_{t+i\delta}(x, \delta) \quad (1)$$

この場合、フレーム間隔とシャッターの開口時間が一致しており、シャッターをフルオープンで撮影した場合のフレームが、これにあたる。これに対し、シャッターの開口時間を $\tilde{m}\delta$ として、時刻 $m\delta \leq t \leq m\delta + \tilde{m}\delta$ の間、部分的にシャッターを開口して撮影したフレームは以下のように表せる。

$$\bar{f}_t(x, \tilde{m}\delta) = \frac{1}{\tilde{m}} \sum_{i=0}^{\tilde{m}-1} f_{t+i\delta}(x, \delta) \quad (2)$$

ここで、 \tilde{m} は $1 \leq \tilde{m} \leq m$ の値をとる。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 (2) に示す低域通過フィルタの影響を考慮する必要がある。

[†]日本電信電話株式会社 NTT サイバースペース研究所

信号 $\bar{f}_t(x, \tilde{m}\delta)$ に対して、区間 L を単位として動き補償 (推定変位置量 \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 における真の変位置量を $d_m(x)$ としている。

$$\begin{aligned} \sigma_e^2 &= \sum_x |\bar{f}_t(x, \tilde{m}\delta) - \bar{f}_{t-m\delta}(x + \hat{d}_m, \tilde{m}\delta)|^2 \\ &= \sum_x |\bar{f}_{t-m\delta}(x + d_m(x), \tilde{m}\delta) \\ &\quad - \bar{f}_{t-m\delta}(x + \hat{d}_m, \tilde{m}\delta) + n(\tilde{m})|^2 \\ &= \sum_x \left| \frac{1}{\tilde{m}} \sum_{i=0}^{\tilde{m}-1} \{f_{t-m\delta+i}(x + d_m(x), \delta) \right. \\ &\quad \left. - f_{t-m\delta+i}(x + \hat{d}_m, \delta)\} + n(\tilde{m}) \right|^2 \\ &= \sum_x \left| \left\{ \frac{\sum_{i=0}^{\tilde{m}-1} \frac{d}{dx} f_{t-m\delta+i}(x, \delta)}{\tilde{m}} \right\} (d_m(x) - \hat{d}_m) \right. \\ &\quad \left. + \phi(x) + n(\tilde{m}) \right|^2 \end{aligned} \quad (3)$$

ここで、 $\phi(x)$ はテイラー展開の2次以降の項である。また、 $n(\tilde{m})$ は撮像デバイスの熱雑音等に起因する雑音成分である。フレーム間での動きを等速運動とみなせば、変位置量はフレーム間隔 (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 (3) に代入し、さらに、テイラー展開の一次近似を利用し、次式を得る。

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

ここで、 $A(m, \tilde{m}), B(m, \tilde{m}), N(m, \tilde{m})$ 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。

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

なお、上式の導出では、雑音成分 n が他の成分と統計的に独立であると仮定して、近似を行っている。また、 $\varepsilon(x) = v_m(x) - \hat{v}_m$ とおいた。

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

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

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

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

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

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。

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

シャッターの開口率 κ を m と \tilde{m} の比として、次式のように定義する。

$$\frac{\tilde{m}}{m} = \kappa (\leq 1)$$

とおく。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 に対する比であることを考慮すると、 \tilde{m} は次式となる。

$$\tilde{m} = \kappa m = \kappa \frac{F_0}{F}$$

上式を用いて、次式を得る。

$$A(m, \tilde{m}) \simeq \beta \frac{2\sigma_s^2(1-\rho)}{\kappa F_0} F \quad (6)$$

同様に、 $B(m, \tilde{m})$ を次式のように近似する。

$$\begin{aligned} B(m, \tilde{m}) &\simeq 2\gamma\phi(x)\sqrt{A(m, \tilde{m})} \\ &= 2\gamma\phi(x)\sqrt{\beta \frac{2\sigma_s^2(1-\rho)}{\kappa F_0} F} \quad (7) \end{aligned}$$

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。

$$n(\tilde{m})^2 = \sum_{x \in L} \frac{n_0^2}{\tilde{m}} = \|L\| \frac{n_0^2}{\kappa F_0} F \quad (8)$$

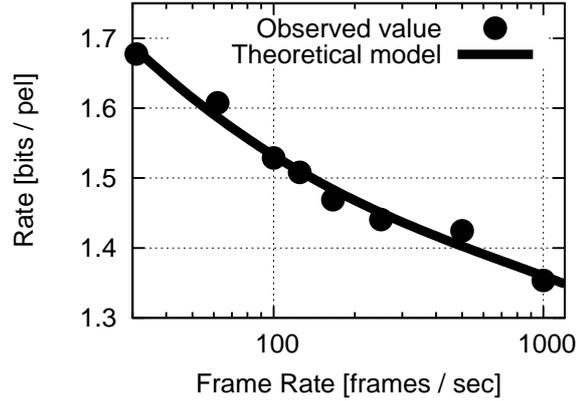
ここで、 $\|L\|$ はセグメント L 内の要素数である。

予測誤差の情報量 $I(F)$ [bits/pel] を次式を用いて見積もる。なお、以下では、対数の底は 2 とする。

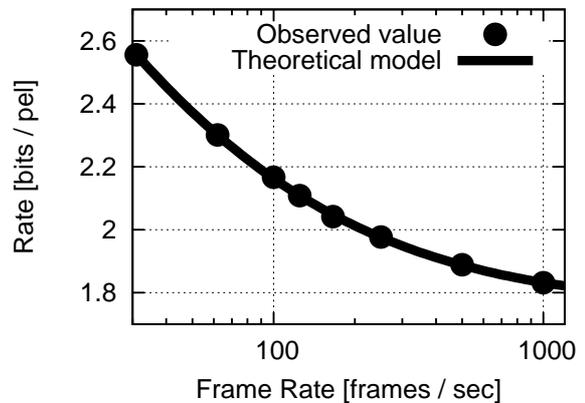
$$\begin{aligned} I(F, \kappa) &= \log(\sigma_e) \\ &\simeq \frac{1}{2} \log\{\hat{A}\kappa^{-1}F^{-1} + \hat{B}\kappa^{-1/2}F^{-1/2} \\ &\quad + \hat{C}\kappa^{-1}F + \hat{D}\} \quad (9) \end{aligned}$$

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

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



(a) Tennis



(b) Golf

図 1: フレームレートと情報量の関係 (開口率 50% の場合)

3 モデルの妥当性の検証

3.1 実画像データとの比較

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。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 (MEM-RECOM 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。このとき、 $\kappa = \frac{\tilde{m}}{m} = 0.5$ とした。また、映像素材はテニス、ゴルフのスイングを撮影したスポーツシーンである。

図1にフレームレートを变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(9)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。動き推定の精度は整数画素精度である。動き推定の評価値はフレーム間予測誤差の二乗和として、フルサーチにより動きベクトルを求めた。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。

4 開口率の最適化

ある特定のフレームレートに対して、フレーム間予測差分信号の情報量を最小化する開口率 κ を求める。 $I(F, \kappa)$ を構成する要素のうち、 \hat{A}, \hat{C}, F は常に正值であるが、 \hat{B} は正負いずれの値も取り得る可能性がある。これは、 \hat{B} に含まれる $\gamma, \phi(x)$ が正負いずれの値も取り得る可能性があるためである。以下では、 \hat{B} が正值、および負値の場合に分けて考える。

\hat{B} が正值の場合、次式を満たす。

$$\frac{dI(F, \kappa)}{d\kappa} < 0$$

つまり、 $I(F, \kappa)$ は κ に関して単調減少関数であり、 κ が取り得る最大値($\kappa = 1$)の場合に、最小となる。 $\kappa = 1$ は、シャッターをフルオープンで撮影したシーケンスに対応する。

一方、 \hat{B} が負値の場合、次式を満たす κ が存在する。

$$\frac{dI(F, \kappa)}{d\kappa} = 0$$

上式の解 κ_0 として、式(9)より次式を得る。

$$\kappa_0 = \frac{4(\hat{A}F^{-1/2} + \hat{C}F^{3/2})^2}{\hat{B}^2} \quad (10)$$

$\kappa = \kappa_0$ のときに、情報量 $I(F, \kappa)$ は最小となる。

上述の最適な開口率を与える式(10)について考察する。式(10)は、 \hat{A} が大きくなれば、符号量を低減するために開口率を大きくする必要があり、さらに、

フレームレートが高い場合は、その必要性が低くなることを示している。 \hat{A} はフレーム間予測誤差に起因する項であり、時間方向の高域成分を含む。時間方向の平滑化処理は、時間方向の低域通過フィルタであるため、時間方向の高域成分の抑圧には有効である。また、フレーム間予測誤差は、フレームレートの増加と共に低減する。前述の \hat{A} と κ_0 の関係は、こうしたフレーム間予測誤差の性質と符合する。また、式(10)は、 \hat{C} が大きくなれば、符号量を低減するために開口率を大きくする必要があり、さらに、フレームレートが高いほど、その必要性が高まることを示している。 \hat{C} は撮像デバイスの熱雑音等の雑音成分に起因する項である。こうした雑音成分の抑圧には、時間方向の平滑化処理が有効である。また、こうした雑音の信号比の電力は、フレームレートが高いほど顕在化する。前述の \hat{C} と κ_0 の関係は、こうした雑音成分の性質と符合する。

5 おわりに

本報告では、フレームレート、開口率、およびフレーム間予測誤差の定量的な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た。さらに、同モデルを用いて、フレーム間差分信号の情報量を最小化する開口率が導出できることを示した。

参考文献

- [1] L.Spillmann et.al., *Visual perception the neurophysiological foundations*. Academic Press, 1990.
- [2] 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.
- [3] 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.