GENERALIZED THEORETICAL MODEL OF RELATIONSHIP BETWEEN FRAME-RATE AND BIT-RATE CONSIDERING LOW PASS FILTERING INDUCED BY SHUTTER OPENING

Yukihiro BANDOH, Kazuya HAYASE, Seishi TAKAMURA, Kazuto KAMIKURA and Yoshiyuki YASHIMA

NTT Cyber Space Laboratories, NTT Corporation
1-1 Hikari-no-oka, Yokosuka, Kanagawa, 239-0847 JAPAN
E-mail : bandou.yukihiro@lab.ntt.co.jp

ABSTRACT

Higher frame-rates are being considered to achieve more realistic representations. Since increasing the frame-rate increases the total amount of information, efficient coding methods are required. However, the statistical properties of such data has not been clarified. This paper establishes, for high frame-rate video, a mathematical model of the relationship between frame-rate and bit-rate. The model incorporates the effect of the low-pass filtering induced by shutter open. By incorporating the open interval of shutter, our model can be extended to describes the various cases of downsampling frame-rates. A coding experiment confirms the validity of the mathematical model.

***Index Terms*—** Video coding, Video signal processing, High frame-rate video

1. INTRODUCTION

Realistic representations using extremely high quality images are becoming increasingly popular. Such representations demand the following four elements: high spatial resolution, large dynamic range, accurate color reproduction, and high temporal resolution. For example, digital cinema[1] and Super High Vision (SHV) [2] offer digital images with high-resolution. Display device and compression method for image with large dynamic range [3] are studied strenuously. Significant efforts to reproduce accurate color are being made by the Natural Vision Project [4].

However, the achievement of high temporal resolution has not been well researched. The current best frame-rate (60 [frames/sec] or [fields/sec]) was simply selected as the lowest rate that well suppressed perceptible flickering. Unfortunately, suppressing flicker is not directly connected to the representation of smooth movement. From the biochemical viewpoint [5], we can estimate that the maximum detectable frame-rate is 150 - 200 [frames/sec].

In order to create more realistic representations, we need to increase the frame-rate. Since high frame-rate video requires more encoded bits than low frame-rate video, the statistical properties of high frame-rate video must be elaborated so as to encode high frame-rate video efficiently. In particular, it is important to have an accurate grasp of the relationship between frame-rate and bit-rate. When the frame-rate increases, the correlation between successive frames increases. It is easily predicted that increasing the frame-rate leads to decrease the encoded bits of inter-frame prediction error. However, the quantitative effect of frame-rate on bit-rate has not been clarified. Conventional studies [6, 7, 8] considered only video frame-rates under 60 [frames/sec], and the statistical properties of high frame-rate video remain unknown. Furthermore, downsampling frame-rate was simply realized by skipping [9].

When high frame-rate video and high-resolution video are taken, you increase the open interval of the shutter in the imaging system in order to retain the input of luminous energy. Increase in the open interval is also effective in reducing jerkiness caused by frame skipping. On the other hand, increase in the open interval causes motion blur because of the integral effect in image sensing device. The integral effect is a kind of low-pass filtering. The low-pass filtering changes the statistical properties of the video signal. The open interval of the shutter can control the trade-off between jerkiness and motion blur which depend on the characteristic of imaging system. Thus, the open interval of the shutter is the important element for considering the relationship between frame-rate and bit-rate. However, the effect of the open interval of the shutter has not been considered in conventional studies.

This paper establishes a mathematical model of the relationship between frame-rate and bit-rate with due consideration of the effect of the low-pass filtering associated with the open interval of the shutter. This model quantifies the impact of frame-rate on bit-rate with variable open interval of the shutter. We verify its validity through encoding experiments.

2. DERIVATION OF MATHEMATICAL MODEL

We derive a mathematical model that describes the relationship between frame-rate and bit-rate. A one-dimensional signal is considered here for simplicity. Let $f_t(x, \delta)$ denote a one-dimensional signal on position x in the t-th frame which was taken with the shutter open in the time interval between t and $t+\delta$. Pixel values in each frame are quantized with 8 [bits] at any interval of shutter open. When the shutter open interval is increased to $m\delta$ (m is a natural number), the corresponding signal $\bar{f}_{mt}(x, m\delta)$ is given by the following equation:

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

In this case, frame interval is the same as the open interval of the shutter. This is the case that the shutter is fully open. When shutter is partly open in the time interval $mt \le t \le mt + \tilde{m}\delta$, taken frame is as follows:

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

where $\tilde{m}$ is in the range $1 \le \tilde{m} \le m$.

When segment L in $\bar{f}_{mt}(x, \tilde{m}\delta)$ is predicted from the previous frame by using estimated displacement ($\hat{d}_m$), the prediction error is given as follows: In the following, $d_m(x)$ is the true displacement at position x.

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

where $\phi(x)$ is the second order remainder term of the Taylor expansion, and $n(m)$ is the noise element. We make the assumption that a moving object exhibits uniform motion across successive frames. This is a likely assumption for a video signal that has a high frame rate. In this case, object displacement is proportional to the frame interval. In other words, the displacement is inversely proportional to frame-rate F. Let $F = (m \cdot \delta)^{-1}$ denote the frame-rate of signal $\bar{f}_{mt}(x, m\delta)$. The displacement error $d_m(x) - \hat{d}_m$ is given by the following equation:

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

where $v_m(x)$ is a constant that depends on position x and $\hat{v}_m$ is a constant that depends on segment L. Henceforth, we substitute $f_t(x)$ for $f_t(x, \delta)$ for simplicity, unless otherwise stated.

By inserting the above equation into equation (3) and using the first order approximation of the Taylor expansion and the assumption that the noise element is statistically independent of the video signal, we obtain:

$$\sigma_e^2 \simeq A(m, \tilde{m})F^{-2} + B(m, \tilde{m})F^{-1} + N(m, \tilde{m}) \quad (4)$$

where $A(m, \tilde{m}), B(m, \tilde{m}), N(m, \tilde{m})$ are as follows:

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

where $\varepsilon(x) = v_m(x) - \hat{v}_m$.

Henceforth, we set $\mu_{\tilde{m}t}(x) = \frac{1}{\tilde{m}}\sum_{i=0}^{\tilde{m}-1} f_{m(t-1)+i}(x)$. $A(m, \tilde{m})$ is expanded as follows:

$$\begin{aligned}
A(m, \tilde{m}) &= \sum_x \left[\varepsilon(x)\frac{d}{dx}\mu_{\tilde{m}t}(x)\right]^2 \\
&\simeq \sum_x [\varepsilon(x)\{\mu_{\tilde{m}t}(x) - \mu_{\tilde{m}t}(x-1)\}]^2 \\
&= \sum_x \{\varepsilon(x)\mu_{\tilde{m}t}(x)\}^2 + \sum_x \{\varepsilon(x)\mu_{\tilde{m}t}(x-1)\}^2 \\
&\quad -2\sum_x \varepsilon(x)^2 \mu_{\tilde{m}t}(x)\mu_{\tilde{m}t}(x-1) \\
&\simeq 2\sum_x \{\varepsilon(x)\mu_{\tilde{m}t}(x)\}^2 \\
&\quad -2\sum_x \varepsilon(x)^2 \mu_{\tilde{m}t}(x)\mu_{\tilde{m}t}(x-1) \\
&\simeq \beta\frac{2\sigma_s^2(1-\rho)}{\tilde{m}^2}\left\{\tilde{m} - \frac{1-\rho}{\rho}\sum_{i>j}\alpha_{i,j}\rho^{|\bar{d}_i - \bar{d}_j|}\right\} \quad (5)
\end{aligned}$$

where, $\beta = \sum_x \varepsilon(x)^2$. In the above approximation, we assume that $\varepsilon(x)$ is statistically independent of $\mu_{mt}(x)$, and we use the following homogeneous model

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

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

and the following approximation

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

where $\bar{d}_i$ and $\bar{d}_j$ are the mean values of $d_i(x)$ and $d_j(x)$ ($x \in L$), respectively, and $\alpha_{i,j}$ is a parameter to collect the approximation of using mean displacement ($\bar{d}_i$ and $\bar{d}_j$).

We can assume that ρ is less than but close to one, since ρ is the autocorrelation coefficient of the image signal. Thus, we have

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

Using this inequality, equation (5) can be approximated as follows:

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

We define the open ratio of shutter (κ) as the ratio of m to $\tilde{m}$ as follows:

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

Since m is the ratio of downsampled frame-rate F to maximum frame-rate F_0, we have

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

Putting above equation into equation (6) gives

$$A(m,\tilde{m}) \simeq \beta \frac{2\sigma_s^2(1-\rho)}{\kappa F_0} F \tag{7}$$

In a similar way, we have

$$\begin{aligned} B(m,\tilde{m}) &\simeq 2\gamma\{\sum_x \phi(x)\}\sqrt{A(m,\tilde{m})} \\ &= 2\gamma\{\sum_x \phi(x)\}\sqrt{\beta \frac{2\sigma_s^2(1-\rho)}{\kappa F_0} F} \end{aligned} \tag{8}$$

where γ is 1 or -1.

Next, let us consider $N(m,\tilde{m})$. Since we assume that the noise element $n(\tilde{m})$ is statistically independent of the video signal, the averaging procedure denoted by equation (1) reduces $n(m)$ as follows:

$$n(\tilde{m})^2 = \sum_{x \in L} \frac{n_0^2}{\tilde{m}} = ||L|| \frac{n_0^2}{\kappa F_0} F \tag{9}$$

where, n_0 is the noise signal included in the sequence at frame-rate F_0, and $||L||$ is the number of elements in segment L .

Assuming that the inter frame prediction error follows a Laplacian distribution and that the differential entropy of the distribution is used, the information of prediction error $I(F,\kappa)$ can be estimated as follows:

$$\begin{aligned} I(F,\kappa) &= \log(\sigma_e) \\ &\simeq \frac{1}{2}\log\{\hat{A}\kappa^{-1}F^{-1} + \hat{B}\kappa^{-1/2}F^{-1/2} + \hat{C}\kappa^{-1}F + \hat{D}\} \end{aligned} \tag{10}$$

where, $\hat{A}$, $\hat{B}$, $\hat{C}$, and $\hat{D}$ are as follows:

$$\hat{A} = \beta \frac{2\sigma_s^2(1-\rho)}{F_0} \tag{11}$$

$$\hat{B} = 2\gamma\{\sum_x \phi(x)\}\sqrt{\beta \frac{2\sigma_s^2(1-\rho)}{F_0}} \tag{12}$$

$$\hat{C} = ||L|| \frac{n_0^2}{F_0} \tag{13}$$

$$\hat{D} = \sum_x \{\phi(x)\}^2 \tag{14}$$

3. CHARACTERISTIC OF DERIVED MODEL

The new feature of our model is to incorporate the parameter (κ) for the open interval of shutter . The parameter κ is in the range of $\frac{1}{m}$ to one. Let us consider the case of $\kappa = \frac{1}{m}$ in which the downsampling frame-rate is realized by frame skip. In this case for $\tilde{m} = 1$, we have $\kappa \frac{F_0}{F} = 1$. Putting this into equation (7) gives as follows:

$$A(m,1) \simeq 2\beta\sigma_s^2(1-\rho) \simeq \beta \left\{ \frac{d}{dx} f_t(x) \right\}^2 \tag{15}$$

In a similar way, we have

$$B(m,1) \simeq 2\gamma\{\sum_x \phi(x)\}\sqrt{\beta \left\{ \frac{d}{dx} f_t(x) \right\}^2} \tag{16}$$

$$n(1)^2 = ||L|| n_0^2 \tag{17}$$

As a result for the case of $\kappa = \frac{1}{m}$, we have

$$\begin{aligned} I(F,\frac{1}{m}) &\simeq \frac{1}{2}\log\{A(m,1)F^{-2} + B(m,1)F^{-1} \\ &\quad + n(1) + \hat{D}\} \end{aligned} \tag{18}$$

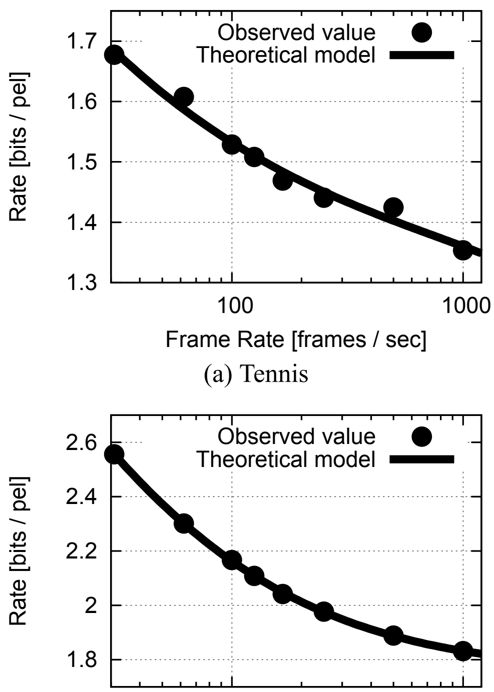
This is consistent with the outcomes of convensional model for the frame-rate controled by frame skip [9]. Thus, the derived model gives the desired expression for the case of $\kappa = \frac{1}{m}$, in spite of approximations introduced in the construction of the model.

4. EXPERIMENTS

4.1. Captured high frame-rate video signal

High frame-rate digital video signals were captured by an NAC MEMRECOM fx RX-3 . The frame-rate was 1000 [frames/sec]. The frame interval equaled the shutter speed of 1/1000 [sec]. The video signal was not gamma corrected. Videos were created of tennis and golf scenes.

In order to identify the relationship between frame-rate and bit-rate, the different frame rates were generated following equation (2) with $\kappa = \frac{\tilde{m}}{m} = 0.5$. The original videos had a resolution of 640 × 480 [pixels]. The original videos in the following experiments consisted of 480 frames at 1000 [frames/sec].



(a) Tennis

(b) Golf

Fig. 1. Relationship between bit-rate and frame-rate

4.2. Verification of model validity

We performed encoding experiments in order to verify the validity of the above-mentioned model. In Figure 1, the dot symbols show the encoding results while the solid lines plot the results of the proposed model. The horizontal axis is the frame-rate [frames/sec] and the vertical axis is the bit-rate [bits / pixel] which is the sum of the entropy of inter frame prediction error and the entropy of motion vectors. The block size used for motion compensation was 16 $\times$ 16 [pixels] in view of the resolution of frame (640 $\times$ 480 [pixels]), instead of 8 $\times$ 8 or 4 $\times$ 4 [pixels]. The accuracy of motion estimation was the quarter pixel; the interpolation filter was taken from MPEG-4. In the inter-frame prediction, the previous frame was used as the reference frame. Parameters $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ were obtained by least-squares estimation.

As shown in Figure 1, the results of the experiments well agree with the values yielded by the proposed model. We observe that the proposed model covers over a broad range of frame-rates from 30 [frames/sec] to 1000 [frames/sec]. We can confirm that our mathematical model well approximates of these experimental results, which supports the validity of the assumptions used in deriving our mathematical model.

5. CONCLUSION

In this paper, we analytically derive a mathematical model that quantifies the relationship between frame-rate and bit-rate for high bit rate videos. Furthermore, we incorporates the effect induced by the open interval of shutter into the model. The incorporation of the open interval generalizes our model so that it can describe the effect of low-pass filtering in case of downsampling frame-rate. Encoding experiments confirm that the model is valid. We believe that the proposed model gives useful information for developing high frame-rate encoding systems. The authors thank referees for careful reading for our manuscript and for giving useful comments.

6. REFERENCES

[1] W.Husak. Economic and other considerations for digital cinema. *Signal Processing: Image Communication*, Vol. 19, No. 9, pp. 921–936, Oct. 2004.

[2] S.Sakaida, K.Iguchi, S.Gohshi, A.Ichigaya, M.Kurozumi, and Y.Nishida. The super hi-vision codec. *ICIP*, 2007. to be appeared in this special session.

[3] R.Mantiuk, G.Kraczyk, K.Myszkowski, and H.Seidel. High dynamic range image and video compression - fidelity. *ICIP*, 2007. to be appeared in this special session.

[4] K. Ohsawa, T. Ajito, H. Fukuda, Y. Komiya, H. Haneishi, M. Yamaguchi, and N. Ohyama. Six-band hdtv camera system for spectrum-based color reproduction. *J. Imaging Science and Technology*, Vol. 48, No. 2, pp. 85–92, 2004.

[5] L.Spillmann and J.S.Werner. *Visual perception the neurophysiological foundations*. Academic Press, 1990.

[6] H. Song and C.-C.J.Kou. Rate control for low-bit-rate video via variable encoding frame rates. *IEEE trans. Circuit Systems and Video Technology*, Vol. 11, No. 4, pp. 512–521, 2001.

[7] Y. Inazumi, T. Yoshida, Y. Sakai, and Y. Horita. Frame rate optimization under a bit rate constraint. *SPIE, Visual Communications and Image Processing*, Vol. 4671, pp. 1130–1142, Jan. 2002.

[8] Y.Shishikui. A study on modeling of the motion compensation prediction error signal. *IEICE Trans. Communications*, Vol. E75-B, No. 5, pp. 368–376, 1992.

[9] Y.Bandoh, S.Takamura, K.Kamikura, and Y.Yashima. Theoretical model of relationship between frame-rate and bit-rate for encoding high frame-rate video signal. *ICIP2005*, Vol. 1, pp. 841–844, 2005.