Rate distortion theory for image and video coding

G. Tziritas

Department of Computer Science, University of Crete, and
Institute of Computer Science, Foundation for Research and Technology-Hellas
P.O. Box 1385, 711 10 Heraklion, Crete, Greece
e-mail : tziritas@csi.forth.gr

Abstract

We give the rate distortion function for three image sequence coders: intra-frame, inter-frame and motion compensating inter-frame. The intensity function distribution is assumed to be Gaussian, and we suppose spatio-temporal separability and an isotropic model for the spatial part of the intensity function. We give quantitative results justifing the use of the above three coding modes in an image sequence coder. Moreover the obtained results show that the hybrid coding (motion compensating prediction + spatial compression) is more interesting as motion analysis becomes more effective.

1 Introduction

Let us consider Figure 1, where prediction is restricted to the preceding frame. We suppose that

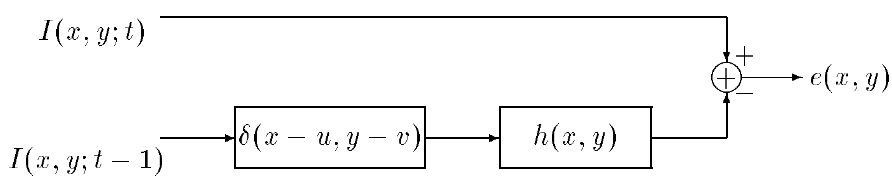

Figure 1: Generic structure of a temporal prediction

the signal $I(x, y; t)$ is stationary, and that the displacement vector (u, v) is a constant parameter, *i.e.*, independent from (x, y). The filter $h(x, y)$ is linear and shift-invariant. We also suppose that the covariance function of $I(x, y; t)$ is separable in time and space. Be it $S_I(f_x, f_y)$ the power spectral density of the spatial part of $I(x, y; t)$. We thus obtain the power spectral density function of $e(x, y)$ as follows

$$S_e(f_x, f_y) = \left(1 + |H(f_x, f_y)|^2 - 2\rho\Re[H(f_x, f_y)]\right) S_I(f_x, f_y) \tag{1}$$

where $H(f_x, f_y)$ is the transfer function of the $h(x, y)$ filter, and ρ is the temporal correlation coefficient of the intensity function.

Several particular cases can be considered within this general formulation. Let us stick to two cases which correspond to inter-frame predictors used in practice. The first one is a simple inter-frame predictor, with $H(f_x, f_y) = \exp(j2\pi(f_x u + f_y v))$. The second one is an inter-frame predictor with motion compensation. We can then write $H(f_x, f_y) = \exp(j2\pi(f_x d_x + f_y d_y))$, where (d_x, d_y) is the estimation error of the displacement vector. In this case $H(f_x, f_y)$ is a random function, because the vector (d_x, d_y) is random.

The structure in Figure 1 does not take into account that the predictor and the quantizer are coupled. Thus the quantization noise on the picture at instant $t - 1$ is neglected. If the distortion is low, this approximate is acceptable. With a high compression ratio, the coding noise is supposed to be spatially and temporally white, and its variance is supposed to equate distortion D, knowing that the coding noise is equal to that of the prediction error quantization. We shall later consider the two cases of inter-frame coding, with or without motion compensation, to establish the relation between bit rate and distortion. In addition, we shall consider firstly intra-frame coding, to illustrate in which conditions inter-frame coding with motion compensation is more efficient.

2 Intra-frame Coding

We shall consider an isotropic model for the spatial covariance of $I(x, y; t)$, as provided by O'Neal and Natarajan (1977) with power spectral density

$$S_I(f_x, f_y) = \frac{f_0 \sigma_I^2}{2\pi(f^2 + f_0^2)^{3/2}} = S_0(f) \tag{2}$$

where $f = \sqrt{f_x^2 + f_y^2}$. We considered that the spatial coordinates were continuous, because in the inter-frame with motion compensation case, an non-integer displacement may occur. But since we are interested in compressing digital pictures, we suppose that the random field is band-limited by an ideal and invariant by rotation low-pass filter. For this model, using Hayes *et al.* [3] formulas for Gaussian random 2-D fields, we find the distortion, as a function of parameter θ

$$\frac{D(\theta)}{\sigma_I^2} = \begin{cases} \pi\theta f^2(\theta) + f_0 \left(\frac{1}{\sqrt{f^2(\theta)+f_0^2}} - \frac{1}{\sqrt{\frac{1}{\pi}+f_0^2}} \right) & \text{if } f(\theta) \leq \frac{1}{\sqrt{\pi}} \\ \theta & \text{if } f(\theta) > \frac{1}{\sqrt{\pi}} \end{cases} \tag{3}$$

and the bit rate

$$R(\theta) = \begin{cases} \frac{3\pi}{4\ln 2} \left(f_0^2 \ln \frac{f_0^2}{f^2(\theta)+f_0^2} + f^2(\theta) \right) & \text{if } f(\theta) \leq \frac{1}{\sqrt{\pi}} \\ \frac{3\pi}{4\ln 2} \left(f_0^2 \ln \frac{f_0^2}{\frac{1}{\pi}+f_0^2} + \frac{1}{\pi} \right) + \frac{1}{2}\log_2 \frac{\theta_0}{\theta} & \text{if } f(\theta) > \frac{1}{\sqrt{\pi}} \end{cases} \tag{4}$$

with $\theta = \frac{f_0}{2\pi\left(f^2(\theta)+f_0^2\right)^{3/2}}$ and $\theta_0 = \frac{f_0}{2\pi\left(\frac{1}{\pi}+f_0^2\right)^{3/2}}$.

For memoryless coding, *i.e.*, with no spatial processing, of this same information source, and always with a mean quadratic criterion, the rate-distortion function is given by [Berger, 1971]

$$R = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma_I^2}{D} & 0 < D \leq \sigma_I^2 \\ 0 & D > \sigma_I^2 \end{cases} \tag{5}$$

3 Inter-frame Coding without Motion Compensation

For the model considered (2) and for memoryless coding the rate distortion function is as follows

$$R = \frac{1}{2}\log_2 \left(\frac{2\sigma_I^2(1 - \rho\exp(-2\pi f_0\sqrt{u^2+v^2}))}{D} + 1 \right) \tag{6}$$

In this case the prediction gain is greater than 1, if $\sqrt{u^2+v^2} < \frac{1}{2\pi f_0}\ln 2\rho$, for $\rho > 0$. This condition clearly shows that inter-frame coding without motion compensation is worthwile, only if displacement is small, relatively to the spatial variations of the picture.

4 Motion Compensating Inter-frame Coding

In this case, we can write [Girod, 1987], with $f_x^2 + f_y^2 \leq 1/\pi$

$$S_e(f_x, f_y) = 2\left(1 - \rho\Phi(f_x, f_y)\right) S_I(f_x, f_y) + D \tag{7}$$

supposing that the probability density function of (d_x, d_y) is symmetric relatively to (0,0), $\Phi(f_x, f_y)$ being the characteristic function of (d_x, d_y). We hence suppose that the probability density function of (d_x, d_y) is isotropic, as given by

$$p(d_x, d_y) = \frac{2}{\pi\sigma_d^2} \exp\left(-\frac{2d}{\sigma_d}\right), \quad d = \sqrt{d_x^2 + d_y^2} \tag{8}$$

and consequently the characteristic function is invariant by rotation. For the isotropic model of covariance used in (2), we can then write

$$R = \pi \int_0^{1/\sqrt{\pi}} f \log_2\left(\frac{2(1 - \rho\Phi_0(f))S_0(f)}{D} + 1\right) df \tag{9}$$

In the case of a memoryless coder, the $R(D)$ function is given by

$$R = \frac{1}{2}\log_2\left(\frac{2\sigma_I^2}{D}\left(1 - \frac{\rho}{(1 + \pi\sigma_d f_0)^2}\right) + 1\right) \tag{10}$$

The prediction gain is greater than 1, if the accuracy of motion estimation is such that $\sigma_d < \frac{\sqrt{2\rho} - 1}{\pi f_0}$ $(\rho > 0)$. It is therefore necessary that the accuracy of the estimator be sufficiently high relatively to spatial variations and the temporal correlation coefficient of $I(x, y; t)$.

It is interesting also to compare on a similar basis the inter-frame coders with no spatial memory, with or without motion compensation. Comparison reveals an improvement of motion compensation, provided that $\sigma_d < \frac{\exp(\pi f_0 \sqrt{u^2 + v^2}) - 1}{\pi f_0}$. Then a sufficient condition is obtained: $\sigma_d < \sqrt{u^2 + v^2}$. Ultimately, the results of these various comparisons lead to the use of several modes of transmission: intra-frame, inter-frame with or without motion compensation, and to the selection of the mode best adapted to the spatiotemporal activity of the image sequence.

5 Results

We give in Figure 2 the bit rate in relation to the signal/distortion ratio, according to the formulas hitherto given for the various coder cases. The signal/distortion ratio is determined, in db, by $\mathrm{SNR} = 10\log_{10}(\sigma_I^2/D)$. The linear curve (a) corresponds to the intra-frame memoryless coder (5). The (b) curve corresponds to the optimal intra-frame coder (3, 4). The (c) curve corresponds to the inter-frame coder without motion compensation and no spatial memory, with a displacement module equal to 2 (6). Curve (d) corresponds to a coder with motion compensation and with no spatial memory (10), estimated with $\sigma_d = \frac{1}{f_s} = \sqrt{\pi}/2$, *i.e.*, with an accuracy approaching one pixel. Curve (e) is of the same type as (d), but has a more accurate estimator, $\sigma_d = \frac{0.1}{f_s}$. Curve (f) (resp. (g)) corresponds to the inter-frame coder with estimated motion compensation with $\sigma_d = \frac{1}{f_s}$ (resp. $\sigma_d = \frac{0.1}{f_s}$) and with spatial memory according to (9). The other two parameters involved in the formulas are set as follows : f_0 is such that $\exp(-2\pi\frac{f_0}{f_s}) = 0.95$ and $\rho = 0.999$. These curves are given for bit rate values from 0 to 2 bits/pel, which is the range that presents

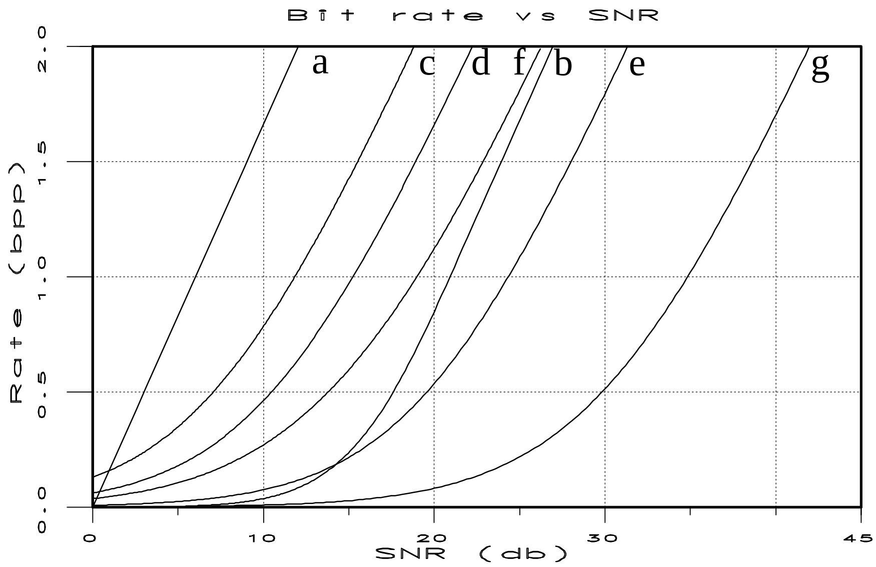

Figure 2: Bit rate according to signal/noise ratio for the various coders

most interest for applications. It is also worth noting that memoryless coders cannot reach bit rate values below 1 bit/pel.

Comparing between curves (d) and (e) on the one hand, and (f) and (g) on the other highlights the importance of displacement estimator accuracy. Note that the gain is higher, by improving the estimator effectiveness, when spatial processing is applied. The usefulness of this processing, which can be carried out either at the signal, or quantizer, or encoder level, is further illustrated by comparing between (d) and (f) on the one hand, and (e) and (g) on the other. The same conclusion, differently expressed, indicates that spatial processing gets more interesting as motion estimation or analysis becomes more effective.

Acknowledgement. The major part of this work was done when the author was with the Laboratoire des Signaux et Systèmes (C.N.R.S., France). A more extensive presentation and demonstrations are given in a book co-authored by C. Labit [5].

References

[1] T. Berger, *Rate distortion theory*, Prentice-Hall, 1971.

[2] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences", *IEEE Jour. Sel. Areas of Communications*, Vol. SAC-5, pp. 1140-1154, Aug. 1987.

[3] J. Hayes, A. Habibi, and P. Wintz, "Rate distortion function for a source model of images", *IEEE Transactions on Information Theory*, Vol. IT-16, pp. 507-509, 1970.

[4] J. O'Neal and T. R. Natarajan, "Coding isotropic images", *IEEE Transactions on Information Theory*, Vol. IT-23, pp. 697-707, Nov. 1977.

[5] G. Tziritas and C. Labit, *Motion analysis for image sequence coding*, Elsevier, 1994.