

The Impact of Aliasing and Quantization on Motion Compensation

Christoph Stiller

INRS-Telecommunications**, 16 Place du Commerce, Verdun, Québec
H3E 1H6 Canada, Tel.:+1 (514) 761-8653, stiller@inrs-telecom.quebec.ca

Abstract. Motion estimation is among the key techniques for a variety of problems in image sequence processing. Even for ideal conditions, when all temporal changes in the image sequence are solely due to motion, the performance of motion compensated prediction is limited by aliasing and quantization noise. This paper provides a theoretical framework for these impact allowing quantification of the power spectral density of the prediction error signal. Both aliasing and quantization are shown to have important impact for typical image sequences. As an illustrative example for an application of the findings, an adaptive in-loop filter for motion compensated prediction is designed. Improvements of up to 1 dB over a standard Wiener filter are demonstrated when the filter adapts on the local motion as suggested by the model calculation.

1 Introduction

Motion information is an important element in analysis, processing, and compression of image sequences. Its basic underlying idea is that some local property of the image is 'nearly' preserved along motion trajectories. Most commonly, the intensity itself is assumed 'almost' constant in this direction. For this reason, most motion estimators include minimization of a norm of the error signal of motion compensated prediction [5].

The knowledge of properties of prediction error images is important for a variety of applications. For motion estimation, statistical models of the prediction error and the motion field allow a Bayesian solution of the estimation problem (e.g., [4, 7]). Motion compensated prediction can be improved by Wiener filtering when the power spectral density of the prediction error image and the original image are given [1]. Hybrid video coding as is employed in current standards (H.261, MPEG-1, MPEG-2) includes transmission of the prediction error image itself. Hence, near rate-distortion optimum encoding implies a good knowledge of the prediction error statistics.

This contribution extends fundamental work of [1, 2] on the analysis of the prediction error by not only accounting for the impact of quantization of motion information but also for the impact of aliasing noise, which is present in any digital imaging system. Experimental studies indicate that the latter impact is far from negligible. On one hand, the analysis gives insight in performance bounds of motion compensated prediction. On the other hand, it provides information

** Major parts of the underlying research were performed at: Institute for Communication Engineering, Aachen University of Technology, Germany

The author recently joined Bosch Research Communications, Hildesheim, Germany

for the design of some elements in video processing, such as motion estimators, motion vector quantizers or in-loop filters and prediction error coders in hybrid video coding.

The rest of this paper is organized as follows. The main part given in the next section is concerned with a theoretical analysis of the impact of aliasing and quantization on motion compensation. Model assumptions for the image sequence and the motion fields allow quantitative evaluation of these impact and give insight into some basic properties. Section 3 provides an illustrative example of an application of these findings for motion compensated prediction as employed in standardized video coding. Conclusions are presented in Section 4.

2 Theoretical Analysis of Motion Compensation

For the analysis, areas of the image sequence are considered, where all intensity changes are solely due to motion, i.e., it is assumed that the image acquisition system is ideal, all objects are diffuse reflecting, illumination is constant and no occlusions are present. Clearly, the performance of motion compensation in such regions will serve as an upper bound for real imagery. A system model for the prediction error is depicted in Figure 1. Input to the system is the intensity

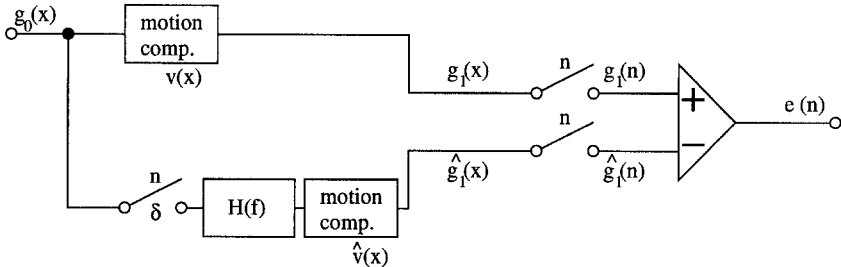


Fig. 1. System model for the prediction error

pattern $g_0(\mathbf{x})$ at time $t=0$ with continuous spatial coordinates $\mathbf{x} = (x, y)^T \in \mathbb{R}^2$. In contrast, $\mathbf{n} = (n, m)^T \in \mathbb{Z}^2$ will denote discrete coordinates of sampled fields. The upper branch of the system models the formation process of the following image $g_1(\mathbf{n})$. Firstly, the pattern $g_0(\mathbf{x})$ is displaced according to some true motion field $\mathbf{v}(\mathbf{x}) = (v(\mathbf{x}), w(\mathbf{x}))^T$ to form the continuous pattern

$$g_1(\mathbf{x}) = g_0(\mathbf{x} - \mathbf{v}(\mathbf{x})) \quad (1)$$

at time $t = 1$. Secondly, it is subsampled to $g_1(\mathbf{n})$ which is the output of the image acquisition system at time $t = 1$. Note that the model assumes existence of a true motion field $\mathbf{v}(\mathbf{x})$.

The motion compensated prediction $\hat{g}_1(\mathbf{n})$ of $g_1(\mathbf{n})$ is composed in the lower branch. A subsampled version of $g_0(\mathbf{x})$ is interpolated by the filter with transfer function $H(\mathbf{f})$, where $\mathbf{f} = (f_x, f_y)^T \in \mathbb{R}^2$ denotes 2D frequency. Subsequently motion compensation employing some estimated motion field $\hat{\mathbf{v}}(\mathbf{x}) = (\hat{v}(\mathbf{x}), \hat{w}(\mathbf{x}))^T$ is performed. Finally, the motion compensated prediction \hat{g}_1 is subsampled onto the image lattice. The prediction error is defined as $e(\mathbf{n}) = g_1(\mathbf{n}) - \hat{g}_1(\mathbf{n})$.

Even for these ideal conditions, motion compensated prediction cannot be expected to perform perfectly, i.e., $e(\mathbf{n}) \neq 0$ for the following two reasons:

- The image sequence is subsampled in both temporal and spatial direction causing motion-dependent aliasing.
- Motion estimation is subject to estimation errors and transmission of motion estimation is subject to quantization noise, i.e., $\hat{\mathbf{v}} \neq \mathbf{v}$.

For a sufficiently small neighborhood, the motion field can be approximated by a first order expansion as

$$\mathbf{v}(\mathbf{x}) \approx (b_x x + c_x, b_y y + c_y)^T.$$

This comprises translation by $\mathbf{c} = (c_x, c_y)^T$ as well as linear dilation by $\mathbf{b} = (b_x, b_y)^T$. If the intensity image and the prediction error are at least weakly stationary, the power spectral density of the latter can be decomposed as

$$\phi_{ee}(\mathbf{f}) = \phi_{g_1 g_1}(\mathbf{f}) + \phi_{\hat{g}_1 \hat{g}_1}(\mathbf{f}) - 2\text{Re}\{\phi_{\hat{g}_1 g_1}(\mathbf{f})\}. \quad (2)$$

For exact estimation of integer $\frac{1}{1+b_x}, \frac{1}{1+b_y}$ the prediction error $e(\mathbf{n})$ is derived to be at least weakly stationary in the appendix. Moreover the power spectral density of the prediction error (2) is shown to become

$$\begin{aligned} \phi_{ee}(\mathbf{f}) = a^2 \mathbb{W}(\mathbf{f}) * \left\{ \frac{1}{a} \phi_{g_0 g_0}(\bar{\mathbf{f}}) + [\phi_{g_0 g_0}(\bar{\mathbf{f}}) * \mathbb{W}(\bar{\mathbf{f}})] \cdot |H(\bar{\mathbf{f}})|^2 \right. \\ \left. - 2\text{Re}\left\{ \phi_{g_0 g_0}(\bar{\mathbf{f}}) e^{-j2\pi \mathbf{c}\mathbf{f}} \left[\left(H^*(\bar{\mathbf{f}}) e^{j2\pi \hat{\mathbf{c}}\mathbf{f}} \right) * \mathbb{W}(\bar{\mathbf{f}}) \right] \right\} \right\} \end{aligned} \quad (3)$$

where $a = (1+b_x)(1+b_y)$, $\bar{\mathbf{f}} = ((1+b_x)f_x, (1+b_y)f_y)^T$ and $\mathbb{W}(\mathbf{f}) = \sum_{\mathbf{n} \in \mathbb{Z}^2} \delta(\mathbf{f} - \mathbf{n})$ denotes the two-dimensional Dirac-pulse sequence (*Shah function*) [6]. This finding gives insight into important properties of prediction error signals as will be discussed in the sequel.

In order to isolate the aliasing impact, a perfect motion estimator is assumed for the moment, i.e., $\hat{\mathbf{b}} = \mathbf{b}$, $\hat{\mathbf{c}} = \mathbf{c}$.

The autocorrelation function of the intensity pattern $g_0(\mathbf{x})$ is modeled isotropic with power 1 [3]

$$\varphi_{g_0 g_0}(\mathbf{x}) = \rho^{\|\mathbf{x}\|},$$

where a typical correlation coefficient is $\rho = 0.9$.

Figure 2 depicts examples of the prediction error energy due to aliasing for an ideal lowpass interpolation filter

$$H(\mathbf{f}) = \begin{cases} 1 & \text{if } |f_x|, |f_y| < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

The aliasing vanishes when all motion vectors connect pixel sites between the two images. It is worth noting that the distance between the original image and the aliasing can be less than 12 dB and hence aliasing is far from negligible in typical coding applications.

When the prediction error due to inaccurate motion vectors is assumed independent from aliasing, it can be derived from eq. (3) as

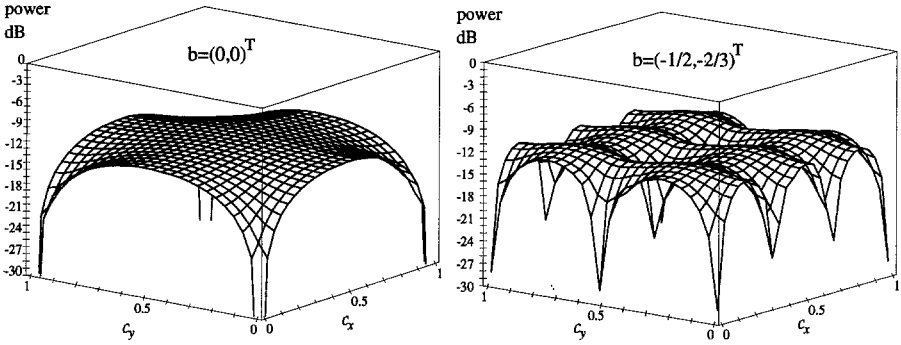


Fig. 2. Prediction error due to aliasing

$$E\{\phi_{ee}(\mathbf{f})\} = a \cdot \mathbb{W}(\mathbf{f}) * \{ \phi_{g_0 g_0}(\bar{\mathbf{f}}) \cdot [1 + |H(\bar{\mathbf{f}})|^2 - 2\text{Re}\{H(\bar{\mathbf{f}}) \cdot F_{\Delta\mathbf{c}}(\mathbf{f})\}] \}, \quad (4)$$

where $F_{\Delta\mathbf{c}}(\mathbf{f}) = E\{e^{j2\pi\Delta\mathbf{c}\mathbf{f}}\}$ is the characteristic function of the error of translatic motion $\Delta\mathbf{c} = \mathbf{c} - \hat{\mathbf{c}}$ denoting the error of translatic motion. For the special case of $\mathbf{b} = \mathbf{0}$, (4) becomes a result of [1] which was derived for rate-distortion optimum coding of gaussian distributed images at high data rate.

When motion inaccuracy is solely caused by linear scalar quantization with small step size Δ , it can be modeled as uniformly distributed with characteristic function $F_{\Delta\mathbf{c}}(\mathbf{f}) = \text{sinc}(\pi\Delta f_x)\text{sinc}(\pi\Delta f_y)$. Figure 3 depicts the expected power of the prediction error according to eq. (4) for this situation.

A straightforward choice for the accuracy Δ in a high quality video coding scheme is to postulate that the error due to motion inaccuracies should at most be of the same magnitude as the error due to aliasing determined by the image acquisition system. As can be seen from Figure 3, this is already satisfied in average for $\Delta = 1$, i.e., pixel accurate motion compensation. A theoretical justification for the common choice of sub-pixel accurate motion vectors can be provided by a closer look at the power spectral density (3) of the prediction error depicted in Figure 4. Visual distortion in the motion compensated prediction occurs mainly in regions with a true motion vector pointing between pixel sites, i.e., $\mathbf{v}(\mathbf{x}) = (n + \frac{1}{2}, m + \frac{1}{2})^T$, where $n, m \in \mathbf{Z}$. In those regions the prediction

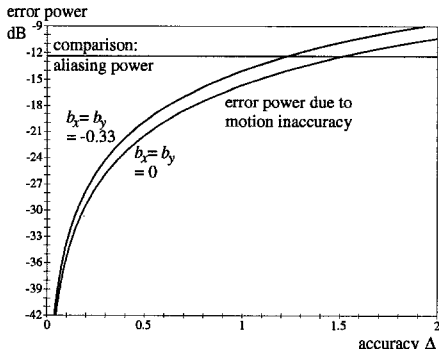


Fig. 3. Prediction error due to inaccurate motion

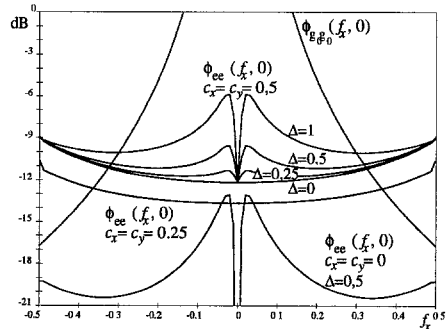


Fig. 4. Power spectral density of prediction error

error is dominated by aliasing at high frequencies. At lower frequencies which are significant to a human observer, an accuracy of $\Delta = \frac{1}{2}$ or even $\Delta = \frac{1}{4}$ is required if no significant distortion is to be added by motion inaccuracy. Hence sub-pixel accuracy is required to reduce the error due to motion inaccuracies to the magnitude of the aliasing error in all regions of the image.

3 An Application – Adaptive Wiener In-Loop Filter for Motion Compensated Prediction

This section sketches an example for an application of the above findings. In video codecs employing motion compensated prediction, a filter with transfer function $W(\mathbf{f})$ is placed behind the motion compensation in the lower branch in Figure 1, known as in-loop filter. Qualitatively it attenuates the prediction at high frequencies. A Wiener filter with transfer function $W_W(\mathbf{f}) = \frac{\phi_{g_{1s}} \hat{g}_{1s}}{\phi_{\hat{g}_{1s}} \hat{g}_{1s}}$ is the fixed filter that achieves optimum SNR, where \hat{g}_{1s} denotes the motion compensated prediction before the filter. A fundamental fact that can be seen from Figure 4 is that the necessary accuracy of motion vectors depends on the true motion. Therefore, the in-loop filter ideally adapts to the local motion. For integer local motion, it should be close to an ideal allpass whereas for non-integer motion it should attenuate the prediction stronger. This suggestion has been verified for a simple block-based motion compensating predictor, as is employed by H.261 and MPEG1/2. Motion estimation was performed by full search block-matching maximizing PSNR of motion compensated prediction.

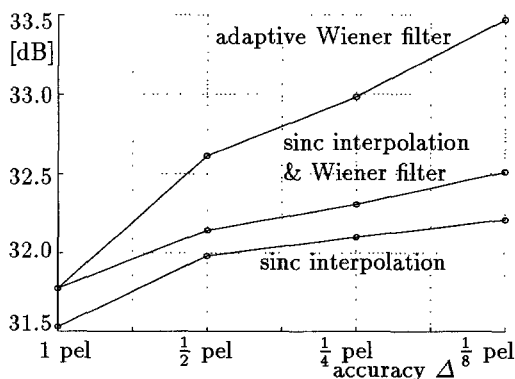


Fig. 5. PSNR of motion compensated prediction for the sequence 'salesman'

Figure 5 depicts the PSNR for the CCITT sequence 'salesman' for adaptive and non-adaptive Wiener in-loop filters, both with 4 tabs. The adaptive Wiener filter was optimized for each individual sub-pixel amount of motion. The improvement of adaptive Wiener filtering over standard Wiener filtering of up to 1 dB confirms the model calculations of the previous section. Since in hybrid video coding, motion information is transmitted to the receiver anyway, these improvements can be achieved without transmission of additional data.

4 Summary and Conclusions

A theoretical framework to investigate the influence of motion vector quantization and aliasing on motion compensated prediction has been presented. Based on a simple model, fundamental properties of the power spectral density of the prediction error signal were examined. It was shown that the impact of aliasing and quantization delimit the performance of motion compensation for typical

situations. Interestingly, these distortions strongly depend on local motion. As an illustrative example for an application of these findings, an adaptive in-loop filter for motion compensated prediction has been designed. Significant improvements over a standard Wiener filter confirm the model calculation. Since in video coding, motion information is transmitted to the receiver anyway, these improvements can be achieved without transmission of additional data.

Appendix

Equation (3) is reached by evaluation of the three terms at the right side of equation (2). From (1) the autocorrelation of g_1 becomes

$$\varphi_{g_1 g_1}(\mathbf{n}) = \varphi_{g_0 g_0} \left(\frac{n}{1+b_x}, \frac{m}{1+b_y} \right)$$

with Fourier transform

$$\phi_{g_1 g_1}(\mathbf{f}) = a \phi_{g_0 g_0}(\bar{\mathbf{f}}) * \mathbb{W}(\mathbf{f}). \quad (5)$$

Without loss of generality the filter H can be assumed a non-ideal lowpass $H(\mathbf{f}) = 0 \forall |\mathbf{f}| > 0.5$. Denoting its autocorrelation function by $\varphi_{hh}^E(\mathbf{x})$ one can write

$$\varphi_{\hat{g}_1 \hat{g}_1}(\mathbf{n}) = \sum_{i,j=-\infty, -\infty}^{\infty, \infty} \varphi_{g_0 g_0}(i, j) \varphi_{hh}^E \left(\frac{n}{1+b_x} - i, \frac{m}{1+b_y} - j \right)$$

with Fourier transform

$$\phi_{\hat{g}_1 \hat{g}_1}(\mathbf{f}) = a^2 \left[(\phi_{g_0 g_0}(\bar{\mathbf{f}}) * \mathbb{W}(\bar{\mathbf{f}})) \cdot |H(\bar{\mathbf{f}})|^2 \right] * \mathbb{W}(\mathbf{f}). \quad (6)$$

Finally, for integer $\frac{1}{1+b_x}, \frac{1}{1+b_y}$ the cross power spectral density $\phi_{\hat{g}_1 g_1}$ exists and after some calculations can be written as

$$\phi_{\hat{g}_1 g_1}(\mathbf{f}) = a^2 \left[\phi_{g_0 g_0}(\bar{\mathbf{f}}) e^{-j2\pi \mathbf{c}\mathbf{f}} \left[(H^*(\bar{\mathbf{f}}) e^{j2\pi \hat{\mathbf{c}}\mathbf{f}}) * \mathbb{W}(\bar{\mathbf{f}}) \right] * \mathbb{W}(\mathbf{f}) \right]. \quad (7)$$

Insertion of (5), (6) and (7) into (2) yields (3).

References

1. B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Sel. Areas Commun.*, vol. 5, pp. 1140–1154, Aug. 1987.
2. B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, pp. 604–612, Apr. 1993.
3. J. Jain and A. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. Commun.*, vol. COM-29, pp. 1799–1808, Dec. 1981.
4. J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-14, pp. 910–927, Sept. 1992.
5. H. Musmann, P. Pirsch, and H.-J. Grallert, "Advances in picture coding," *Proc. IEEE*, vol. 73, pp. 523–548, Apr. 1985.
6. C. Stiller, *Modellbasierte Bewegungsschätzung in Bildfolgen*. PhD thesis, Aachen University of Techn., Fortschr.-Ber., Ser. 10, No. 320, Düsseldorf: VDI-Verlag, 1994.
7. C. Stiller and B. Hürtgen, "Combined displacement estimation and segmentation in image sequences," in *Proc. SPIE/EUROPTO Video Commun. and PACS for Medical Appl.*, vol. 1977, pp. 276–287, Apr. 1993.