Efficiency Analysis of Multihypothesis Motion-Compensated Prediction for Video Coding

Bernd Girod, *Fellow, IEEE*

***Abstract*—Overlapped block motion compensation or B-frames are examples of multihypothesis motion compensation where several motion-compensated signals are superimposed to reduce the bit-rate of a video codec. This paper extends the wide-sense stationary theory of motion-compensated prediction (MCP) for hybrid video codecs to multihypothesis motion compensation. The power spectrum of the prediction error is related to the displacement error probability density functions (pdf's) of an arbitrary number of hypotheses in a closed-form expression. We then study the influence of motion compensation accuracy on the efficiency of multihypothesis motion compensation as well as the influence of the residual noise level and the gain from optimal combination of N hypotheses. For the noise-free limiting case, doubling the number of (equally good) hypotheses can yield a gain of up to $1/2$ bits/sample, while doubling the accuracy of motion compensation (such as going from integer-pel to $1/2$-pel accuracy) can additionally reduce the bit-rate by up to 1 bit/sample independent of N. For realistic noise levels, it is shown that the introduction of B-frames or overlapped block motion compensation can provide larger gains than doubling motion compensation accuracy. Spatial filtering of the motion-compensated candidate signals becomes less important if more hypotheses are combined. The critical accuracy beyond which the gain due to more accurate motion compensation is small moves to larger displacement error variances with increasing noise and increasing number of hypotheses N. Hence, sub-pel accurate motion compensation becomes less important with multihypothesis MCP. The theoretical insights are confirmed by experimental results for overlapped block motion compensation, B-frames, and multiframe motion-compensated prediction with up to eight hypotheses from ten previous frames.**

***Index Terms*—B-frame, hybrid coding, motion compensation, multiframe prediction, multihypothesis motion-compensated prediction, overlapped block motion compensation, sub-pel accuracy, video compression.**

I. Introduction

MOTION-COMPENSATED coding schemes achieve compression by exploiting the similarities between successive frames of a video signal. Often, with such schemes, motion-compensated prediction (MCP) is combined with intraframe encoding of the prediction error employing an 8 × 8 discrete cosine transform. Successful applications range from digital video broadcasting at several megabytes per second down to bit-rates as low as 10 kbps for videophones or Internet video-on-demand applications. Several standards, such as ITU-T Recommendations H.261 [1] and H.263 [2], [3], or ISO MPEG-1 and MPEG-2 [4] are based on this scheme. The new MPEG-4 standard follows the same approach [5], [6].

Most of the work for the design and optimization of video codecs is carried out experimentally. A theoretical treatment of motion-compensated video coding requires many assumptions and simplifications for the analysis of a complicated system processing real-world signals. Nevertheless, even an approximate theory can provide useful insights in the underlying mechanisms and give guidance for the design of state-of-the-art video codecs. A good theoretical framework leads motion-compensated video coding away from heuristics and toward an engineering science.

In 1987, the first comprehensive rate-distortion analysis of MCP was presented [7]. It relates the power spectral density $\Phi_{ee}(\omega_x, \omega_y)$ of the prediction error to the accuracy of motion compensation captured by the probability density function (pdf) of the displacement error. The fundamental equation derived in [7] is

$$\begin{aligned}\Phi_{ee}(\omega_x,\omega_y) = \Phi_{ss}(\omega_x,\omega_y)\cdot(1+|F(\omega_x,\omega_y)|^2 \\ - 2\Re(F(\omega_x,\omega_y)P(\omega_x,\omega_y))) \\ + \Phi_{nn}(\omega_x,\omega_y)|F(\omega_x,\omega_y)|^2 \end{aligned} \quad (1)$$

where

ω_x	horizontal frequency;
ω_y	vertical frequency;
$\Phi_{ss}(\omega_x,\omega_y)$	spatial power spectrum of the input video signal;
$F(\omega_x,\omega_y)$	frequency response of the "loop filter";
$P(\omega_x,\omega_y)$	two-dimensional (2-D) Fourier transform of the displacement error pdf;
$\Phi_{nn}(\omega_x,\omega_y)$	power spectrum of residual noise that cannot be predicted by motion compensation;
$\Re(\cdot)$	real part of a complex number.

The fundamental equation (1) captures the effect that even inaccurate motion compensation still works well for the low spatial frequency components of the signal. Low frequency components do not vary rapidly. For high spatial frequency components, however, a very good accuracy is required since a small offset can lead to a 180° phase shift and thus an increase instead of a reduction of the prediction error. Therefore, the loop filter $F(\omega_x, \omega_y)$ should appropriately attenuate high frequency

Manuscript received March 24, 1997; revised July 14, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Roland Wilson.

The author is with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: girod@ee.stanford.edu).

Publisher Item Identifier S 1057-7149(00)01161-1.

components and effectively switch off motion compensation for high frequencies.

Based on (1), it is shown in [7] that with integer-pel accuracy of the displacement estimate the additional gain by MCP over optimum intraframe encoding is limited to ~0.8 bits/sample in moving areas. Larger gains require fractional-pel accuracy. For $1/2$-pel accuracy, as included in MPEG and in H.263, the gain is limited to ~1.8 bits/sample. Also, the theory explains why a loop filter is essential for good compression performance. An optimum loop filter $F(\omega_x, \omega_y)$ can be derived from (1), resulting in a minimum of the prediction error spectrum of

$$\Phi_{ee}(\omega_x,\omega_y)=\Phi_{ss}(\omega_x,\omega_y) \cdot\left(1-|P(\omega_x,\omega_y)|^2\frac{\Phi_{ss}(\omega_x,\omega_y)}{\Phi_{ss}(\omega_x,\omega_y)+\Phi_{nn}(\omega_x,\omega_y)}\right) \quad (2)$$

Shortly afterwards, the theory was complemented and confirmed by experimental results [8]. In particular, it was shown that signal components that do not obey the paradigm of a piecewise constant translation limit the performance of MCP. It was found that, for a motion compensation block size of 16 × 16 and typical broadcast TV signals, $1/4$-pel accuracy appears to be sufficient, while for videophone signals $1/2$-pel accuracy is desirable. For videophone signals, bilinear interpolation was found to perform almost as well as the best Wiener spatial interpolation/prediction filter, and an additional loop filter is not required. These results also appeared in a journal paper [9] and are summarized in [10]. Recent textbooks discuss motion compensation based on (1) and (2) or simplifications of it, e.g., [11], [12]. Similar analyses have been carried out by Ribas-Corbera and Neuhoff [14]–[17]. In particular, they have considered the rate-constrained motion compensation problem in detail that was introduced in [18] and [19]. Vandendorpe *et al.* have extended the power spectrum versus displacement accuracy analysis to interlaced video [13].

Many codecs today employ more than one motion-compensated prediction signal simultaneously to predict the current frame. The term "multihypothesis motion compensation" has been coined for this approach [20]. A linear combination of multiple prediction hypotheses is formed to arrive at the actual prediction signal. Examples are the combination of past and future frames to predict B-frames in the MPEG or H.263 coding schemes [2]–[4], or the combination of three motion-compensated signals employing "remote motion vectors" in the "Advanced Prediction Mode" of ITU-T Recommendation H.263 [2], [3]. Both schemes have been experimentally shown to yield a significant coding gain over the classical "single-hypothesis" motion compensation [3], [20]–[25]. While theoretical motivations for multihypothesis motion compensation have been presented [20], [24], its rate-distortion efficiency in terms of motion compensation accuracy and number of hypotheses employed has not yet been analyzed. It is therefore the goal of this paper to extend (1) and (2) to multihypothesis motion-compensated prediction, to compute performance bounds and to compare these to the established performance bounds for classical single-hypothesis motion compensation. Section II introduces two performance measures for motion-compensated hybrid coders. Section III reviews the results that are needed to treat multihypothesis motion compensation as a linear prediction problem. Section IV introduces the power spectral density model for inaccurate motion compensation which is numerically evaluated in Section V. Section VI finally compares the theory to established experimental results for overlapped block motion compensation, B-frames, and multiframe motion-compensated prediction.

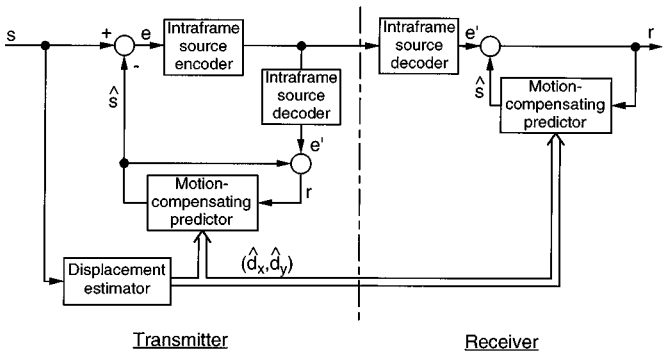

Fig. 1. Block diagram of an MCP hybrid coding scheme.

II. Performance Measures for Motion-Compensated Hybrid Coders

A motion-compensated hybrid coder combines differential pulse code modulation along an estimated motion trajectory of the picture contents with intraframe encoding of the prediction error e (Fig. 1). The displacement estimate $(\hat{d}_x, \hat{d}_y)$ is transmitted in addition to the intraframe-encoded prediction error. At the receiver, the intraframe source decoder generates the reconstructed prediction error e', which differs from e by some reconstruction error. The transmitter contains a replication of the receiver in order to generate the same prediction value $\hat{s}$.

It has been pointed out by several authors that the motion-compensated prediction error signal e is only weakly correlated spatially, e.g., [7]–[12], [26]–[28]. Thus, the potential for redundancy reduction in the intraframe source encoder is relatively small. This finding suggests that the prediction error variance

$$\sigma_e^2 = E\{e^2\} - E^2\{e\} \quad (3)$$

is a useful measure that is related to the minimum achievable transmission bit-rate for a given signal-to-noise ratio [29]. In (3), $E\{\cdot\}$ is the expectation operator. The minimization of prediction error variance (3) is widely used to obtain the displacement vector and control the coding mode in practical systems. A more refined measure is the rate difference

$$\Delta R = \frac{1}{8\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\log_2\left(\frac{\Phi_{ee}(\omega_x,\omega_y)}{\Phi_{ss}(\omega_x,\omega_y)}\right)d\omega_x\,d\omega_y. \quad (4)$$

In (4), $\Phi_{ee}(\omega_x,\omega_y)$ and $\Phi_{ss}(\omega_x,\omega_y)$ are the power spectral densities of the prediction error e and the signal s, respectively. Unlike (3), the rate difference (4) takes the spatial correlation (or spectral flatness) of the prediction error e and the original signal s into account. It represents the maximum bit-rate reduction (in bits/sample) possible by optimum encoding of the prediction error e, compared to optimum intraframe encoding of the signal s for Gaussian wide-sense stationary signals for the same mean

squared reconstruction error [29]. A negative ΔR corresponds to a reduced bit-rate compared to optimum intraframe coding, while a positive ΔR is a bit-rate increase due to motion compensation, as it can occur for inaccurate motion compensation. The maximum bit-rate reduction can be fully realized at high bit-rates, while for low bit-rates the actual gain is smaller [7]. Note that we neglect the rate required for transmitting the displacement estimate $(\hat{d}_x, \hat{d}_y)$ in addition to the prediction error e. The optimum balance between rates for the prediction error signal and displacement vectors strongly depends on the total bit-rate, as discussed, e.g., in [19]. For high rates, it is justified to neglect the rate for the displacement vectors, while for low rates it is essential to take it into account. Throughout this paper, we shall employ ΔR as our performance measure.

III. Multihypothesis Motion Compensation as a Linear Prediction Problem

Let $s[x,y]$ be a scalar 2-D signal sampled on an orthogonal grid with horizontal spacing X and vertical spacing Y. Let $c[x,y]$ be a vector-valued signal (column vector of length N) sampled at the same positions. For the problem of multihypothesis motion compensation, we interpret c as the vector of multiple motion-compensated frames available for prediction, and s as the current frame to be predicted.

Assume that s and c are generated by a jointly wide-sense stationary random process with the real-valued scalar power spectral density $\Phi_{ss}(\omega_x,\omega_y)$, the $N\times N$ power spectral density matrix $\Phi_{cc}(\omega_x,\omega_y)$, and the $N\times 1$ cross spectral density vector $\Phi_{cs}(\omega_x,\omega_y)$. Power spectra and cross spectra are defined according to

$$\Phi_{ab}(\omega_x,\omega_y)=\mathcal{F}_*\{E\{a[x+x',y+y']b^H[x,y]\}\} \quad (5)$$

where

a and b	complex column vectors;
b^H	transposed complex conjugate of b;
$[x,y]\in\Pi$	sampling locations;
$E\{a[x+x',y+y']b^H[x,y]\}$	matrix of space-discrete cross correlation functions between the components of a and b which (for wide-sense stationary random processes) does not depend on x and y but only on the relative horizontal and vertical shifts x' and y';
$\mathcal{F}_*\{\cdot\}$	2-D band-limited discrete-space Fourier transform shown in (6).

$$\mathcal{F}_*\{\cdot\}=\sum_{[x,y]\in\Pi}(\cdot)e^{-j\omega_x\frac{x}{X}-j\omega_y\frac{y}{Y}} \quad \forall|\omega_x|<\pi,\quad |\omega_y|<\pi. \quad (6)$$

As in [9], we do not require the origin $(x=0,y=0)$ to coincide with one of the samples. Thus, (6) is slightly more general than the conventional definition of the 2π-periodic discrete-space Fourier transform, e.g. see [30]. We restrict the region of support of the Fourier transform to the baseband and do not consider baseband replications. This restriction greatly simplifies dealing with fractional-pel shifts in the following without sacrificing generality.

It is well-understood how to predict a scalar signal s from the vector-valued signal c, such that the mean square of the prediction error

$$e[x,y]=s[x,y]-f[x,y]*c[x,y] \quad (7)$$

is minimized. Nevertheless, we present a brief summary here to have the important results handy. In (7), the asterisk $*$ denotes generalized 2-D convolution, i.e., $f[x,y]*c[x,y]=\sum_{[u,v]\in\Pi_f} f[u,v]\cdot c[x-u,y-v]$, where the result is calculated for all values $[x,y]\in\{[x,y]:[x-u,y-v]\in\Pi_c\wedge[u,v]\in\Pi_f\}$, with Π_f and Π_c the sampling grids of $f[x,y]$ and $c[x,y]$, respectively. $f[x,y]$ is a row vector of impulse responses. The power spectral density of the prediction error is

$$\begin{aligned}\Phi_{ee}(\omega_x,\omega_y)=\Phi_{ss}(\omega_x,\omega_y)&-\Phi_{sc}(\omega_x,\omega_y)F^H(\omega_x,\omega_y)\\&-F(\omega_x,\omega_y)\Phi_{cs}(\omega_x,\omega_y)\\&+F(\omega_x,\omega_y)\Phi_{cc}(\omega_x,\omega_y)F^H(\omega_x,\omega_y). \quad (8)\end{aligned}$$

In (8), $F(\omega_x,\omega_y)=\mathcal{F}_*\{f[x,y]\}$ is a row vector of N complex transfer functions. We shall omit the independent variables (ω_x,ω_y), when there is no danger of confusion. The above equation thus can be written more compactly as

$$\begin{aligned}\Phi_{ee}&=\Phi_{ss}-\Phi_{sc}F^H-F\Phi_{cs}+\\&\quad F\Phi_{cc}F^H=\Phi_{ss}-2\Re\{F\Phi_{cs}\}+F\Phi_{cc}F^H. \quad (9)\end{aligned}$$

Note that in (9), Φ_{ee} and Φ_{ss} are real-valued, F and Φ_{cs} are complex row and column vectors, respectively, and Φ_{cc} is a positive definite $N\times N$ matrix.

The prediction error power spectrum Φ_{ee} is minimized separately at each frequency by the optimum transfer function

$$F=\Phi_{cs}^H\Phi_{cc}^{-1}, \quad (10)$$

as can be verified by inserting $F=\Phi_{cs}^H\Phi_{cc}^{-1}+\Delta F$ into (9) and observing that Φ_{ee} increases by $\Delta F\Phi_{cc}\Delta F^H\geq 0$ when F deviates from the global minimum (10).

The corresponding minimum prediction error power spectral density is found by inserting (10) into (8) or (9)

$$\Phi_{ee}=\Phi_{ss}-\Phi_{cs}^H\Phi_{cc}^{-1}\Phi_{cs}. \quad (11)$$

Since the optimum multiple input filter F minimizes the prediction error power spectrum Φ_{ee} separately at each spatial frequency (ω_x,ω_y), it simultaneously minimizes the prediction error variance (3) and the rate difference ΔR (4).

IV. Power Spectral Models for Inaccurate Motion Compensation

Since we are interested in performance bounds of multihypothesis motion compensation, we shall assume that optimum filters F according to (10) are used and (11) holds. Then, the only remaining problem is an appropriate statistical model of

c and s that yields Φ_{ss}, $\Phi_{cs} = \Phi_{sc}^H$, and Φ_{cc}. As in [7] and [9]–[11], we assume that an image v possesses an isotropic spatial power spectrum

$$\Phi_{vv}(\omega_x, \omega_y) = \frac{2\pi}{\omega_0^2} \cdot \left(1 + \frac{\omega_x^2 + \omega_y^2}{\omega_0^2}\right)^{-\frac{3}{2}}. \quad (12)$$

The power spectrum is normalized to an overall signal variance $\sigma_v^2 = 1$. It corresponds to an isotropic exponentially decaying autocorrelation function. ω_0 is a parameter that captures the correlation between adjacent pixels. For the numerical results, we shall set ω_0 to correspond to an average correlation factor of 0.93 that can be measured between horizontally or vertically adjacent pels in a typical video signal. We now assume that an individual frame s of a video sequence is a noisy, shifted version of v, such that its power spectrum is

$$\begin{aligned}\Phi_{ss}(\omega_x, \omega_y) &= \Phi_{vv}(\omega_x, \omega_y) + \Phi_{nn0}(\omega_x, \omega_y) \\ &= \frac{2\pi}{\omega_0^2} \cdot \left(1 + \frac{\omega_x^2 + \omega_y^2}{\omega_0^2}\right)^{-\frac{3}{2}} + \Phi_{nn0}(\omega_x, \omega_y). \quad (13)\end{aligned}$$

We will come back to the noise n_0 with power spectrum $\Phi_{nn0}(\omega_x, \omega_y)$ in the following discussion. For now, it suffices to say that n_0 is typically white noise with a variance $\sigma_n^2 \ll 1$.

Obviously, multihypothesis motion-compensated prediction should work best if we compensate the true displacement of the scene exactly for each candidate prediction signal. Less accurate compensation will degrade the performance. However, even for exact motion compensation, there will be residual signal components that are present in one frame, but not in the other.

To capture the limited accuracy of motion compensation, we associate a displacement error $(\Delta_{xi}, \Delta_{yi})$ with the ith component $c_i[x, y]$ of the vector of motion-compensated candidate signals $c[x, y]$. The horizontal displacement error Δ_{xi} is normalized relative to the horizontal sampling interval X, the vertical displacement error Δ_{yi} relative to Y. Further, we assume that the "clean" video signal v can be predicted up to some residual noise n_i from $c_i[x, y]$, if its associated displacement error would vanish. Fig. 2 illustrates this model. Since the current frame $s = v + n_0$ contains additional noise n_0 uncorrelated from v, the "noisy" video signal s can be predicted up to residual $n_0 + n_i$ from c_i, if the displacement error $(\Delta_{xi}, \Delta_{yi}) = (0, 0)$.

The displacement error reflects the inaccuracy of the displacement vector used for the motion compensation. Even the best displacement estimator will never be able to measure the displacement vector field without error. More fundamentally, the displacement vector field can never be completely accurate since it has to be transmitted as side information with a limited bit-rate. The "noise" $n_i + n_0$ comprises all signal components that cannot be described by a translatory displacement model. This includes not only camera noise and quantization noise due to source coding of the signal c, but also illumination changes, resolution changes due to zoom and varying distance between camera and object, sampling artifacts, and so on. We deliberately split up the noise into separate components, n_i with power spectral density $\Phi_{nni}(\omega_x, \omega_y)$ and n_0 with power spectrum $\Phi_{nn0}(\omega_x, \omega_y)$. In this fashion, we can model signal components n_i that are associated with c_i, but statistically independent from s, as well as signal components n_0 that are associated with s, but independent from each of the c_i.

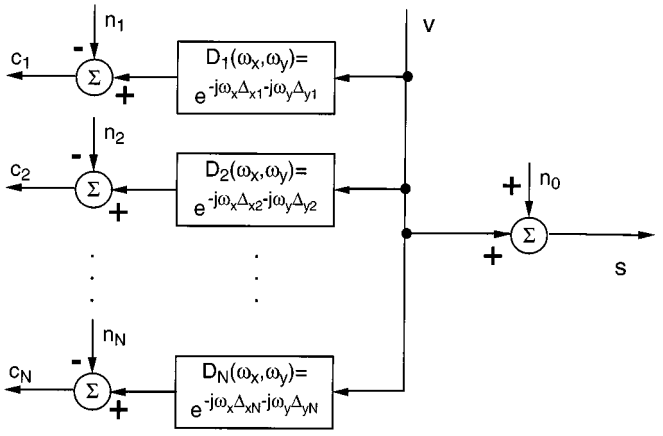

Fig. 2. Statistical model of multihypothesis motion-compensated prediction.

Assuming that all the $n_i, i = 0, \ldots, N$ are uncorrelated with signal v, and that n_0 is uncorrelated with $n_i, i = 1, \ldots, N$, the model in Fig. 2 yields

$$\Phi_{cs} = \Phi_{cv} = D\Phi_{vv} \quad (14)$$

and

$$\Phi_{cc} = D\Phi_{vv}D^H + \Phi_{nn} \quad (15)$$

with the abbreviation

$$D = \begin{pmatrix} e^{-j\omega_x\Delta_{x1} - j\omega_y\Delta_{y1}} & & \\ & e^{-j\omega_x\Delta_{x2} - j\omega_y\Delta_{y2}} & \\ & \vdots & \\ & e^{-j\omega_x\Delta_{xi} - j\omega_y\Delta_{yi}} & \\ & \vdots & \\ & e^{-j\omega_x\Delta_{xN} - j\omega_y\Delta_{yN}} & \end{pmatrix}. \quad (16)$$

$\Phi_{nn}(\omega_x, \omega_y)$ is an $N \times N$ diagonal matrix with elements $\Phi_{nni}(\omega_x, \omega_y), i = 1, \ldots, N$ if the individual noise components n_i are uncorrelated which shall be assumed in the following.

We now interpret the displacement errors $(\Delta_{xi}, \Delta_{yi}), i = 1, \ldots, N$ as random variables which are statistically independent from v and $n_i, i = 0, \ldots, N$. With that, we rewrite (14) and (15) as

$$\Phi_{cs} = E\{D\}\Phi_{vv} \quad (17)$$

and

$$\Phi_{cc} = E\{D\Phi_{vv}D^H\} + \Phi_{nn} = \Phi_{vv}E\{DD^H\} + \Phi_{nn}. \quad (18)$$

We observe that (see (19) at the bottom of the next page). Thus, the ith component $P_i(\omega_x, \omega_y)$ of $E\{D\}$ is the 2-D Fourier transform $\mathcal{F}\{p_i(\Delta_{xi}, \Delta_{yi})\}$ of the continuous 2-D pdf of the displacement error Δ_{xi}, Δ_{yi}. Since the integral of a proper pdf is one, $P_i(0, 0) = 1$. Toward higher frequencies, $P_i(\omega_x, \omega_y)$ decays quickly for inaccurate motion compensation and slowly for

accurate motion compensation. For the expected value in (18), we obtain

$$E\{DD^H\} = \begin{pmatrix} 1 & P_1P_2^* & P_1P_3^* & \cdots & P_1P_N^* \\ P_2P_1^* & 1 & P_2P_3^* & \cdots & P_2P_N^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_NP_1^* & P_NP_2^* & P_NP_3^* & \cdots & 1 \end{pmatrix}. \tag{20}$$

Equation (20) holds under the assumption that the displacement errors Δ_{xi}, Δ_{yi} and Δ_{xj}, Δ_{yj} are mutually statistically independent for $i \neq j$. Note that we do not require that the individual horizontal and vertical components Δ_{xi} and Δ_{yi} are independent. Combining (11), (17)–(19), and (20) we obtain the fundamental equation of multihypothesis MCP (see (21) at the bottom of the page) with

$$\alpha_i = \frac{\Phi_{nni}}{\Phi_{vv}}. \tag{22}$$

Knowing the pdf's of the displacement errors $(\Delta_{xi}, \Delta_{yi})$ and the spectral noise-to-signal power ratios $\alpha_i(\omega_x, \omega_y)$, we can use (21) and (4) to calculate the maximum rate difference due to multihypothesis motion compensation.

If we do not use the optimum filters (10), but some other suboptimum transfer function F, we can combine (9), (17)–(19), and (20) to obtain (23), shown at the bottom of the page. Note that for $N = 1$ and $\alpha_0 = 0$, we obtain the fundamental equations of motion-compensated prediction (1) and (2) as special cases of (23) and (21).

In analogy to the single-hypothesis case discussed in the introduction, (21) and (23) capture the effect that motion compensation is easy for low spatial frequency components of the video signal but difficult for high spatial frequencies. Since high-frequency components change rapidly, a high motion compensation accuracy is required. If motion compensation is inaccurate, then $P_i \ll 1 \forall i$ in (21) and (23) for high frequencies, and $\Phi_{ee}/\Phi_{ss} \approx 1$ results in (21), and, assuming additionally that $\alpha_i = \alpha_j \forall i, j$, then $\Phi_{ee}/\Phi_{ss} \approx 1 + FF^H$ in (23). Obviously, the optimum filter F is a low pass filter that removes high frequency components from c that are too noisy or that change too rapidly for the given displacement error.

V. Numerical Results

Let us now use the results in Section IV to evaluate some interesting cases numerically. [7] studies the case of a flat noise power spectrum

$$\Phi_{nn}(\omega_x, \omega_y) = \sigma_n^2 \tag{24}$$

and single-hypothesis motion-compensated prediction with an isotropic Gaussian displacement error pdf of variance σ_Δ^2

$$p(\Delta_x, \Delta_y) = \frac{1}{2\pi\sigma_\Delta^2} \exp\left(-\frac{\Delta_x^2 + \Delta_y^2}{2\sigma_\Delta^2}\right). \tag{25}$$

Can we do better if we combine several hypotheses even though they have the same displacement pdf (25) and noise spectrum (24)? Fig. 3–5 show the rate difference (4) as a function of

$$E\{D\} = \int \cdots \int_{-\infty}^{\infty} p(\Delta_{x1}, \Delta_{y1}, \ldots, \Delta_{xN}, \Delta_{yN}) \begin{pmatrix} e^{-j\omega_x\Delta_{x1} - j\omega_y\Delta_{y1}} \\ e^{-j\omega_x\Delta_{x2} - j\omega_y\Delta_{y2}} \\ \vdots \\ e^{-j\omega_x\Delta_{xi} - j\omega_y\Delta_{yi}} \\ \vdots \\ e^{-j\omega_x\Delta_{xN} - j\omega_y\Delta_{yN}} \end{pmatrix} d\Delta_{x1}, \Delta_{y1} \ldots d\Delta_{xN}, d\Delta_{yN}$$
$$= \begin{pmatrix} \mathcal{F}\{p_1(\Delta_{x1}, \Delta_{y1})\} \\ \mathcal{F}\{p_2(\Delta_{x2}, \Delta_{y2})\} \\ \vdots \\ \mathcal{F}\{p_N(\Delta_{xN}, \Delta_{yN})\} \end{pmatrix} =: \begin{pmatrix} P_1(\omega_x, \omega_y) \\ P_2(\omega_x, \omega_y) \\ \vdots \\ P_N(\omega_x, \omega_y) \end{pmatrix} \tag{19}$$

$$\frac{\Phi_{ee}}{\Phi_{ss}} = 1 - (1 + \alpha_0)^{-1} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix}^H \begin{pmatrix} 1 + \alpha_1 & P_1P_2^* & P_1P_3^* & \cdots & P_1P_N^* \\ P_2P_1^* & 1 + \alpha_2 & P_2P_3^* & \cdots & P_2P_N^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_NP_1^* & P_NP_2^* & P_NP_3^* & \cdots & 1 + \alpha_N \end{pmatrix}^{-1} \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} \tag{21}$$

$$\frac{\Phi_{ee}}{\Phi_{ss}} = 1 + (1 + \alpha_0)^{-1} \left(-2\Re\left\{ F \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} \right\} + F \begin{pmatrix} 1 + \alpha_1 & P_1P_2^* & P_1P_3^* & \cdots & P_1P_N^* \\ P_2P_1^* & 1 + \alpha_2 & P_2P_3^* & \cdots & P_2P_N^* \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_NP_1^* & P_NP_2^* & P_NP_3^* & \cdots & 1 + \alpha_N \end{pmatrix} F^H \right) \tag{23}$$

displacement error variance for different residual noise levels RNL $= 10\log_{10}(\sigma_n^2)$ dB. Before interpreting these curves, a comment on the horizontal axis calibration is in order.

The horizontal axes in Figs. 3–5 are calibrated by $\log_2(\sqrt{12}\sigma_\Delta)$ to support an easier interpretation of the diagrams. Consider a perfect displacement estimator that always estimates the true displacement. Then, the displacement error (Δ_x, Δ_y) is entirely due to rounding. In moving areas with sufficient variation of motion, the displacement error is uniformly distributed between $\pm 2^{\beta-1}$ and $\pm 2^{\beta-1}$ where $\beta = 0$ for integer-pel accuracy, $\beta = -1$ for $1/2$-pel accuracy, $\beta = -2$ for $1/4$-pel accuracy, etc. The minimum displacement error variance in moving areas is

$$\sigma^2_{\Delta/\min} = \frac{2^{2\beta}}{12}. \tag{26}$$

It turns out that the precise shape of the displacement error pdf has hardly any influence on the variance of the motion-compensated prediction error, σ_e^2, as long as the displacement error variance σ_Δ^2 does not change. A uniform pdf and a Gaussian pdf yield essentially the same variances σ_e^2. Thus, for a given $\beta, \log_2(\sqrt{12}\sigma_\Delta)$ cannot be smaller than β in moving areas. On the other hand, a good displacement estimator can probably come close to that value. Note that this requires more sophisticated motion compensation than the blockwise constant displacement common today [1], [2]. Buschmann [31] shows that for typical CIF videoconferencing sequences and blockwise constant displacement, an additional displacement error variance of about 10% of the displacement variance σ_d^2 is introduced for 16×16 blocks, and of 5% of σ_d^2 for 8×8 blocks. For example, for 16×16 blocks, he measures displacement error variances that correspond to $\log_2(\sqrt{12}\sigma_\Delta) = 0.57, \log_2(\sqrt{12}\sigma_\Delta) = 0.52$, and $\log_2(\sqrt{12}\sigma_\Delta) = 0.51$ for integer-pel, half-pel, and quarter-pel accuracy, respectively, (as opposed to the theoretical $\log_2(\sqrt{12}\sigma_\Delta) = 0, -1, -2$).

A. Noise-Free Case

Fig. 3 shows the rate difference (4) as a function of $\log_2(\sqrt{12}\sigma_\Delta)$ for the practically noise-free case with RNL $= -60$ dB. Note that we should avoid setting $\sigma_n^2 = 0$ because we then cannot invert the matrix in (21) at $\omega_x = \omega_y = 0$. For $N = 1$, Fig. 3 shows again the known result that the gain due to integer-pel accurate motion compensation is limited to $\sim$0.8 bits/sample [7]. For $1/2$-pel accuracy, the gain is limited to $\sim$1.8 bits/sample. For each refinement of the accuracy by a factor of 2, the bit-rate decreases by about 1 bit/sample. This also holds for the multihypothesis curves $N > 1$.

Doubling the number of hypotheses decreases the bit-rate by $1/2$ bits/sample in the part of the diagram, where the curves in Fig. 3 are straight and parallel. Thus, quadrupling the number of hypotheses provides as much gain as refining the displacement accuracy horizontally and vertically by a factor of two for the noise-free case. Note that we can also interpret a refinement of the resolution of the displacement vector from integer-pel to $1/2$-pel, or from $1/2$-pel to $1/4$-pel as quadrupling the number of hypotheses for motion compensation. For example, for the

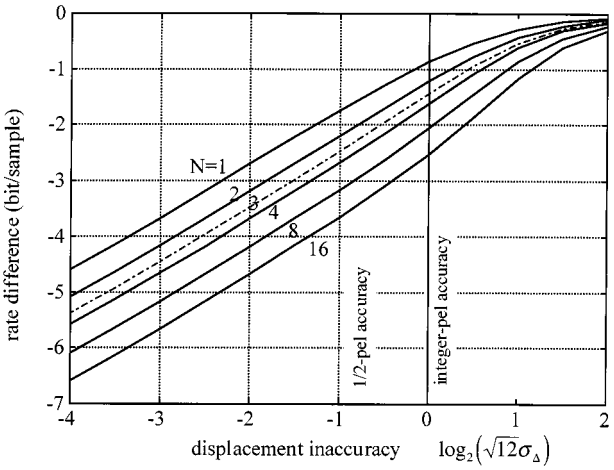

Fig. 3. Rate difference compared to optimum intraframe coding due to multihypothesis motion-compensated prediction as a function of displacement error variance for combining different numbers of hypotheses N. Residual noise level RNL $= -60$ dB.

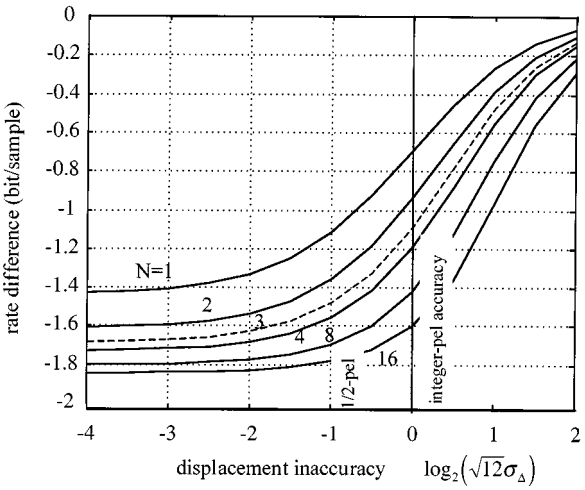

Fig. 4. Rate difference compared to optimum intraframe coding due to multihypothesis motion-compensated prediction as a function of displacement error variance for combining different numbers of hypotheses N. Residual noise level RNL $= -24$ dB.

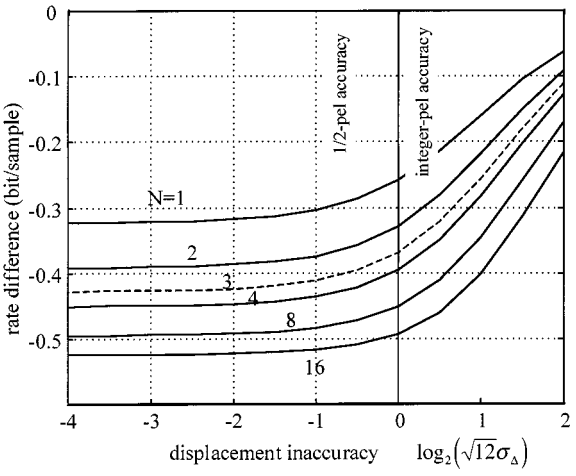

Fig. 5. Rate difference compared to optimum intraframe coding due to multihypothesis motion-compensated prediction as a function of displacement error variance for combining different numbers of hypotheses N. Residual noise level RNL $= -12$ dB.

refinement from integer-pel to $1/2$-pel, we obtain three additional polyphase representations of the same image ($1/2$ pel to the right, $1/2$ line up, $1/2$ pel to the right and $1/2$ line up), each with integer-pel resolution.

At $1/2$-pel or integer-pel accuracy, the curves in Fig. 3 are no longer straight and the rate differences become somewhat smaller. E.g., going from $N = 1$ to $N = 2$ hypotheses decreases the bit-rate by 0.3 bits/sample at integer-pel accuracy.

B. Influence of Residual Noise

Fig. 3 suggests that almost arbitrary bit-rate savings are possible by using more and more accurate motion compensation. This would indeed be the case if the hypothesis signals were noise-free. More realistic numerical results for the range of accurate motion compensation are obtained by taking into account the noise components $n_i, i = 0, \ldots, N$. Figs. 4 and 5 illustrate the efficiency of single-hypothesis and multihypothesis motion compensation as a function of displacement error variance for residual noise levels RNL $= -24$ dB and RNL $= -12$ dB. The observations that were reported in [9], [10] for single-hypothesis motion compensation can be extended to multihypothesis motion compensation as well. Beyond a certain "critical accuracy," the possibility of further improving prediction by more accurate motion compensation is small. The critical point is at a low displacement error variance for low noise variances and at a high displacement error variance for high noise variances. Doubling the number of hypotheses reduces the effect of residual noise by up to $1/2$ bits/sample for the noise-free case, but the gain is usually much smaller with noise. For example, when going from $N = 1$ to $N = 2$ at RNL $= -12$ dB, the gain is less than 0.1 bits/sample. This is due to the fact that the noise power spectra $\Phi_{nni}, i = 1, \ldots, N$ are significantly larger than Φ_{vv} for all but the lowest frequencies and most of the spectrum is suppressed by the optimum filter F. When combining more hypotheses, the independent noise components $n_i, i = 1, \ldots, N$ are more effectively suppressed, such that more of the spectrum is recovered for motion compensated prediction. Ultimately, for small displacement error variance σ_Δ^2 and large $N, \Phi_{ee} = \Phi_{nn0}$, i.e., the noise component n_0 associated with the current frame cannot be reduced by prediction. Therefore, we observe diminishing returns and ultimately a saturation for increasing N. Because of the combination of these diminishing returns and more efficient prediction with increasing N for larger σ_Δ^2, the critical accuracy moves to larger displacement error variances with increasing N. For example, for RNL $= -12$ dB, half-pel accuracy is required to reach a rate within 0.03 bits/sample of the lowest rate for $N = 1$, while for $N = 16$, this is the case already for integer-pel accuracy. This implies that accurate motion compensation is less important with a multihypothesis scheme.

If we again estimate the maximum gain possible by introduction of B-frames instead of P-frames, but now for the high-noise case RNL $= -12$ dB and $1/2$-pel accuracy, we can read a difference of 0.07 bits/sample between $N = 1$ and $N = 2$ from Fig. 5. This is a significantly smaller gain (only about $1/7$) than for the noise-free case (Fig. 3). Interestingly, increasing the number of hypotheses from $N = 1$ to $N = 4$ is more effective than increasing the accuracy from integer-pel to $1/2$-pel for the high noise case RNL $= -12$ dB and also for RNL $= -24$ dB. Only for the academic case RNL $= -60$ dB is there a slight advantage of going from integer-pel to $1/2$-pel accuracy over combining four integer-pel motion hypotheses. For the high noise case RNL $= -12$ dB, even an increase from $N = 1$ to $N = 2$ is more effective than increasing the accuracy from integer-pel to $1/2$-pel. For practically interesting cases, we conclude that combining two predictions in B-frames or utilizing OBMC as described in [24] or as practiced in the H.263 Advanced Prediction Mode [2], [3] can yield as good or better a performance increase than the refinement from integer to $1/2$-pel accuracy. Of course, in an efficient codec, we should combine both.

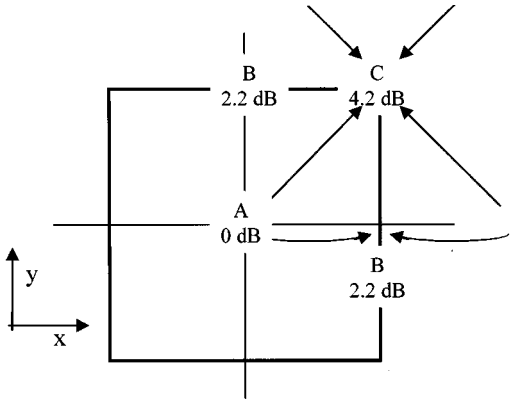

Fig. 6. Performance estimate for overlapped block motion compensation. At point A, $N = 1$ hypothesis is used; at point B, $N = 2$. At the block corner C, $N = 4$. The gains given are relative to single hypothesis MCP for a residual noise level RNL $= -24$ dB.

C. Averaging Hypotheses

So far, we studied the case when an optimum filter (10) is used. How important is that? Can we "get away" with simply averaging the N hypotheses and not filtering spatially? This case with $F = (1/N \ 1/N \ \ldots \ 1/N)$ is shown in Figs. 6–8. The dashed lines correspond to averaging the hypotheses while the solid lines correspond to optimum filtering according to (10). For the noise-free case RNL $= -60$ dB, we can observe a gain due to spatial filtering only for inaccurate motion compensation. However, with the more realistic RNL $= -24$ dB (Fig. 7) and RNL $= -12$ dB (Fig. 8), spatial filtering becomes increasingly important. Interestingly, spatial filtering is less important if the number of hypotheses N increases. For example, in Fig. 7, about 0.13 bits/sample are lost if spatial filtering is omitted for $N = 1$ and integer-pel accuracy, while this loss is negligible for $N = 4$ and $N = 8$. This is because averaging several hypotheses reduces the noise n such that higher frequency components can benefit from motion-compensated prediction if the displacement error variance is small enough. Note that for inaccurate motion compensation, the bit-rate required for the prediction error e can actually be higher than for the original signal s if spatial filtering is omitted. Nevertheless, averaging several equally good hypotheses always reduces the bit-rate. When comparing Figs. 6–8, we can also conclude that in a practical codec (with reasonably accurate motion compensation), the major purpose of an optimum F will typically be

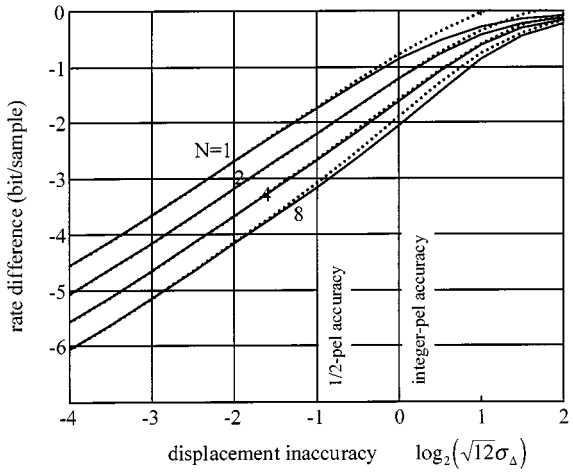

Fig. 7. Rate difference compared to optimum intraframe coding due to multihypothesis motion-compensated prediction as a function of displacement error variance for combining different numbers of hypotheses N. Residual noise level RNL $= -60$ dB. Solid lines assume an optimum filter F, dashed curves show the case where hypotheses are simply averaged.

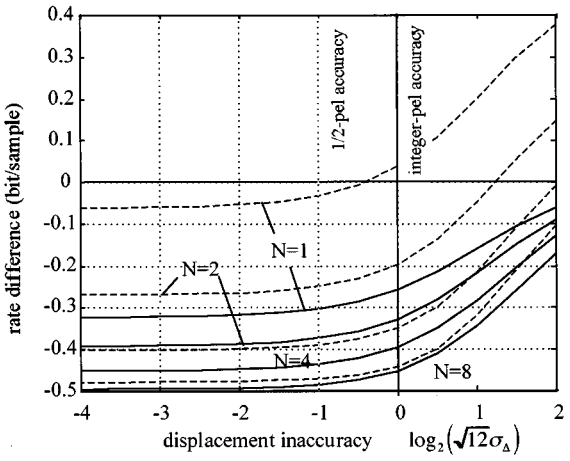

Fig. 9. Rate difference compared to optimum intraframe coding due to multihypothesis motion-compensated prediction as a function of displacement error variance for combining different numbers of hypotheses N. Residual noise level RNL $= -12$ dB. Solid lines assume an optimum filter F, dashed curves show the case where hypotheses are simply averaged.

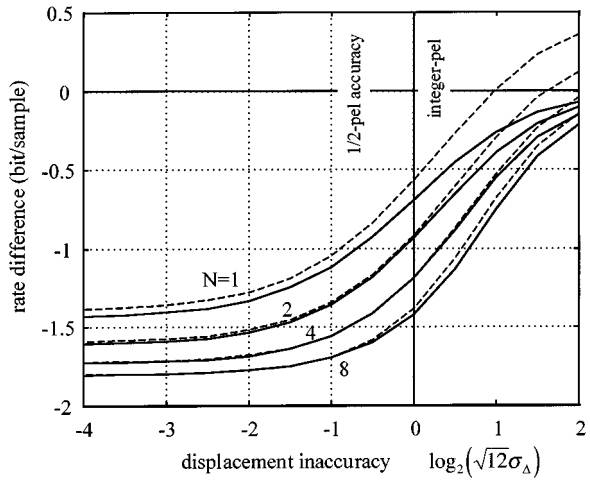

Fig. 8. Rate difference compared to optimum intraframe coding due to multihypothesis motion-compensated prediction as a function of displacement error variance for combining different numbers of hypotheses N. Residual noise level RNL $= -24$ dB. Solid lines assume an optimum filter F, dashed curves show the case where hypotheses are simply averaged.

that of noise reduction, while the gain by taking into account the displacement error pdf's is relatively small (Fig. 6).

VI. Comparison with Experimental Results

A. Overlapped Block Motion Compensation

We can use the curves in Figs. 6–8 to obtain an insight into the performance of overlapped block motion compensation (OBMC) as, for example, described by Orchard and Sullivan [24]. In their work, 16 × 16 tessalating blocks are extended to 32 × 32 windows with a 2:1 overlap horizontally and vertically. Thus, at each pixel there are four distinct motion-compensated versions of the previous frame, i.e., four hypotheses. The windows taper off, i.e., a linear combination of these four hypotheses is formed with spatially slowly varying weights, such that in the center of a block only one hypothesis is used, in the middle between two horizontally or vertically adjacent blocks two hypotheses are averaged, while at the corner in between four blocks, four hypotheses are averaged. In their experiments, Orchard and Sullivan carried out motion compensation with integer-pel accuracy, and we assume that the line $\log_2(\sqrt{12}\sigma_\Delta) = 0$ applies. Since the dashed curves in Fig. 7 are approximately equidistant around that line, refining the estimate by incorporating Buschmann's results [31] (see Section V) does not change the resulting numbers significantly. Orchard and Sullivan did not use the rate difference measure (4)

TABLE I
Variances of the Motion-Compensated Prediction Error for Three Different Sequences Using Forward Prediction, Backward Prediction, and Bidirectional Prediction from the Previous and the Successive Frame for Blocksizes 16 × 16 and 8 × 8. Values Are Stated in dB as PSNR $= 10\log_{10} 255^2/\sigma_e^2$

Blocksize 16 × 16			
Signal/accuracy	Forward prediction	Backward prediction	Bidirectional prediction
Salesman, integer-pel	34.75	34.71	36.23
Salesman, 1/2-pel	35.38	35.36	36.56
Flowergarden, integer-pel	26.40	26.70	30.73
Flowergarden, 1/2-pel	27.73	28.08	30.70
Kiel Harbour, integer-pel	31.61	30.50	34.73
Kiel Harbour, 1/2 pel	33.96	32.79	35.05
Blocksize 8 × 8			
Signal/accuracy	Forward prediction	Backward prediction	Bidirectional prediction
Salesman, integer-pel	35.99	35.94	37.19
Salesman, 1/2-pel	36.74	36.70	37.61
Flowergarden, integer-pel	27.47	27.74	31.92
Flowergarden, 1/2-pel	29.10	29.42	31.92
Kiel Harbour, integer-pel	31.80	30.70	35.52
Kiel Harbour, 1/2 pel	34.18	33.02	35.53

for performance evaluation, but prediction error variance (3). To compare the numbers, we simply convert between (3) and (4) assuming that $\Delta R = 1$ bits/sample corresponds to 6.02 dB in prediction error variance. This is justified since the prediction error spectrum is basically flat in all the cases compared. We use the numbers for RNL $= -24$ dB (Fig. 7) to illustrate the argument in Fig. 9. At point A in the middle of a block, there is no gain by overlapped neighboring blocks, since only a single hypothesis is used. In the middle of the edges (points B), two estimates are combined based on the displacement vectors of the adjacent blocks with roughly equal accuracy, thus the case $N = 2$ with a gain of 2.2 dB applies. At the block corners, four displacement vectors from neighboring blocks are combined with roughly equal accuracy, hence the case $N = 4$ with a gain of about 4.2 dB applies. This is consistent with the experimental observation reported in [24] that OBMC improves prediction most effectively at the edges and especially in the corners of a block. The overall gain results from a combination of the spatially varying gain due to OBMC and can be estimated to be around $(0 + 2.2 + 2.2 + 4.2)/4$ dB ≈ 2.1 dB by averaging the above figures in an interpolation argument. Orchard and Sullivan measured a best performance for "standard" OBMC (i.e., without state variable conditioning) of 1.3 dB (1.7 dB within the training set). At the block corners, where our model computes a maximum gain of 4.2 dB, [24] reports 3.2 dB (within the training set). Considering that the numbers are strongly dependent on the choice of the parameter RNL, the performance figures predicted by our model calculations are encouragingly close to the experimental results reported in [24]. If we repeat the OBMC performance estimate for RNL $= -12$ dB (Fig. 8), we estimate a smaller gain of 1.3 dB.

B. B-Frames

We now compare the numerical results of our analysis with experimental results for unidirectionally predicted P-frames and bidirectionally predicted B-frames. The prediction error variances in Table I were obtained by averaging over ten luminance frames of the video sequences *Salesman*, *Flowergarden*, and *Kiel Harbour* which were processed in the noninterlaced Common Intermediate Format (352 pels × 288 lines, 30 fps). Motion compensation uses a block size of 16 × 16 or 8 × 8 pels without block overlap. For half-pel accuracy, bilinear interpolation is used. Forward prediction uses only the previous (original) frame for prediction, whereas backward prediction uses the following (original) frame. Bidirectional prediction simply averages the forward and backward prediction signals. The motion estimator uses an exhaustive search in a 16 × 16 search window, half-pel displacements are obtained by refinement of the best integer-pel displacement vector. We discuss the results obtained for blocksize 16 × 16 in the following, the findings for blocksize 8 × 8 are similar.

For Salesman, the gain obtained by $1/2$-pel accuracy over integer-pel accuracy is about 0.6 dB for both forward and backward prediction. The gain by using bidirectional prediction for integer-pel accuracy is more than twice as large. This confirms the insight obtained from the model calculations that increasing the number of hypotheses from $N = 1$ to $N = 2$ can be more effective than increasing the accuracy from integer-pel to $1/2$ pel for sufficiently high RNL. Similar observations are made for the Flowergarden and the Kiel Harbour sequences. For Flowergarden, $1/2$-pel accuracy yields a gain of 1.3 and 1.4 dB for forward and backward prediction with integer-pel accuracy, respectively, while bidirectional prediction yields more than 4 dB improvement. Compared to the Salesman sequence, frame-to-frame changes in the Flowergarden sequence can be modeled more accurately by locally constant displacements from frame to frame, hence the relative gains are larger. The prediction error variance values for Salesman include the stationary background, hence the overall PSNR values are greater than for Flowergarden, where the entire picture is moving. Kiel Harbour is a sequence particularly suitable for motion compensation, since its motion only consists of a zoom. A $1/2$-pel accuracy gains about 2.3 dB, while bidirectional prediction gains more than 3 dB for integer-pel accuracy. As predicted by the theory, the gain by $1/2$-pel accuracy over integer-pel accuracy is much smaller when combined with bidirectional prediction in all cases. In one case, we even measure a minor loss when combining both. In general, the relative gains observed in our experiments are well within the range of the model calculations.

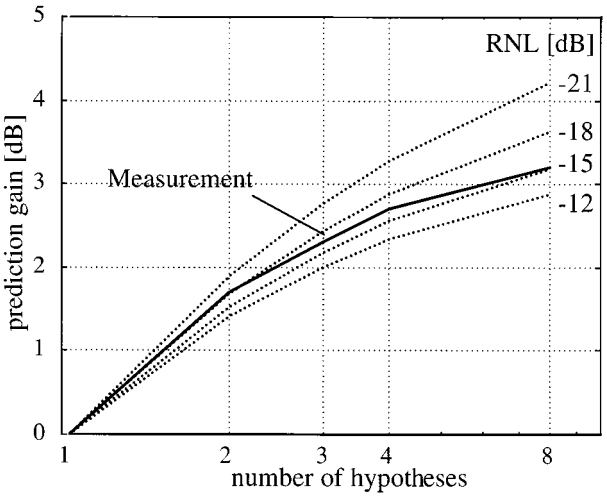

Fig. 10. Prediction gain for integer-pel accuracy of motion compensation measured for $N = 2, 3, 4, 8$ hypotheses over single hypothesis prediction [32]. The dashed lines are model calculations for different residual noise levels RNL $= -12 \ldots -21$ dB.

C. Multiframe Prediction

Finally, we compare the theoretical results to a multiframe motion-compensated prediction method that has been presented in more details in [32]. The motion-compensated block-based predictor searches in up to ten previous frames for an optimal combination of N hypotheses. Fig. 10 shows the gains in prediction error variance relative to single-hypothesis motion-compensated prediction with integer-pel accuracy for the Foreman sequence (QCIF resolution, 7.5 fps, 10 s). A blocksize of 16 × 16 was used, hypotheses are simply averaged. The multihypothesis predictor gains 1.7 dB averaging $N = 2$ hypotheses, and more than 3 dB, if $N = 8$ hypotheses are combined. Fig. 10

also shows the theoretical predictions using (23) for different residual noise levels RNL. The findings reported in [32] are consistent with our theory.

VII. Conclusion

In this paper, we have extended the wide-sense stationary theory of motion-compensated prediction to multihypothesis motion compensation. The power spectrum of the prediction error is related to the displacement error pdf's of an arbitrary number of the hypotheses and a vector of residual noise spectra that captures the components of the motion-compensated hypothesis signal that do not obey the paradigm of translatory motion. The theory can be used to study the influence of motion compensation accuracy on the efficiency of multihypothesis motion compensation as well as the influence of the residual noise level and the gain from optimal combination of N hypotheses. Several important conclusions can be drawn from a numerical evaluation of the theory, some of which have already been reported in previous experimental studies, while others are new.

- An optimum combination of N hypotheses always lowers the bitrate for increasing N. If each hypothesis is equally good in terms of displacement error pdf, doubling N can yield a gain of 0.5 bits/sample if there is no residual noise.
- Doubling the accuracy of motion compensation, such as going from integer-pel to $1/2$-pel accuracy, can reduce the bitrate by up to 1 bit/sample independent of N for the noise-free case.
- If realistic residual noise levels are taken into account, the gains possible by doubling the number of hypotheses, N, decreases with increasing N. We observe diminishing returns and, ultimately, saturation.
- If the power of residual noise components increases, quadrupling and ultimately doubling the number of hypotheses N becomes more effective than doubling the accuracy of motion compensation. As a consequence, the introduction of B-frames or overlapped block motion compensation can provide a larger gain than an increase from integer-pel to $1/2$-pel accuracy.
- The critical accuracy beyond which the gain due to more accurate motion compensation is small moves to larger displacement error variances with increasing noise and increasing number of hypotheses N. Hence, sub-pel accurate motion compensation becomes less important with multihypothesis MCP.
- Spatial filtering of the motion-compensated candidate signals becomes less important if more hypotheses are combined.

In order to make the problem of multihypothesis motion compensated prediction analytically tractable, we had to make several simplifying assumptions, such as the stationarity and the mutual independence of several of the random variables involved, the spatial constancy of the displacement error, the Gaussian statistics of the video signal itself, or the high bitrate and the optimal performance of the encoder. Also, we neglected the rate for transmitting displacement vectors. Mostly, these assumptions make the gain obtainable by motion-compensated coding larger rather than smaller, and we usually interpret the theoretical results as performance limits of a practical coder. Experimental results obtained with actual video sequences often show significantly smaller gains, especially for low bitrate coding. Nevertheless, the theoretical analysis can isolate the various effects that determine the efficiency of multihypothesis motion-compensated prediction and thus provide insight into successful algorithms like OBMC or B-frames and guidance for new multihypothesis schemes.

Acknowledgment

The author gratefully acknowledges the important contributions by Dr. U. Horn, T. Wiegand, and M. Flierl, who were keen critics of the theory and provided the experimental results for single- and multihypothesis prediction. Prof. R. M. Gray had valuable suggestions on optimum multivariate linear prediction. Insightful comments and suggestions by Dr. Y. Wang and by the anonymous reviewers helped to significantly improve the presentation of this material.

References

[1] *Video Codec for Audiovisual Services at p×64 kbit/s*, 1990. ITU-T Rec. H.261.
[2] *Video Coding for Narrow Telecommunications Channels at <64 kbit/s*, 1996. ITU-T Rec. H.263.
[3] B. Girod, E. Steinbach, and N. Färber, "Performance of the H.263 video compression standard," *J. VLSI Signal Process.*, no. 17, pp. 101–111, 1997.
[4] *Generic Coding of Moving Pictures and Associated Audio Information, Part 2: Video*, Mar. 1994. ISO/IEC 13 818 (ITU-T H.262).
[5] *Signal Process.: Image Commun.*, vol. 9, no. 4, May 1997. Special Issue on MPEG-4, Part 1.
[6] *Signal Process. Image Commun.*, vol. 10, July 1997.
[7] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1140–1154, Aug. 1987.
[8] B. Girod and F. Joubert, "Motion-compensating prediction with fractional pel accuracy for 64 kbit/s coding of moving video," in *Proc. Int. Workshop on 64 kbit/s Coding of Moving Video*, Hannover, Germany, June 1998, pp. 1.1.1–1.1.2.
[9] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Trans. Commun.*, vol. 41, pp. 604–612, Apr. 1993.
[10] ——, "Motion compensation: Visual aspects, accuracy, and limitations," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds. Norwell, MA: Kluwer, 1993, pp. 125–152.
[11] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards—Algorithms and Architectures*. Norwell, MA: Kluwer, 1995.
[12] J.-R. Ohm, *Digitale Bildcodierung*. Berlin, Germany: Springer-Verlag, 1995.
[13] L. Vandendorpe, L. Cuvelier, and B. Maison, "Statistical properties of prediction error images in motion compensated interlaced image coding," in *Proc. ICIP-95*, vol. 3, Washington, DC, Oct. 1995, pp. 192–195.
[14] J. Ribas-Corbera and D. L. Neuhoff, "Optimal bit allocations for lossless video coders: Motion vectors vs. difference frames," in *Proc. ICIP-95*, Washington, D.C., Oct. 1995, pp. 180–183.
[15] J. Ribas-Corbera and D. L. Neuhoff, "On the optimal motion vector accuracy for block-based motion-compensated video coders," in *Proc. SPIE Dig. Video Compr.*, San Jose, CA, Jan./Feb. 1996, pp. 302–314.
[16] J. Ribas-Corbera and D. L. Neuhoff, "Reducing rate/complexity in video coding by motion estimation with block adaptive accuracy," in *Proc. Visual Communication Image Processing VCIP'96*, Orlando, FL, Mar. 1996, pp. 615–624.
[17] J. Ribas-Corbera and D. L. Neuhoff, "On the optimal block size for block-based, motion-compensated video coders," in *Conf. Visual Communication Image Processing, (VCIP'97)*, San Jose, CA, Jan.–Feb. 1997.

[18] G. J. Sullivan and R. L. Baker, "Rate-distortion optimized motion compensation for video compression using fixed and variable size blocks," in *Proc. GLOBECOM*, Nov. 1991, pp. 85–90.
[19] B. Girod, "Rate-constrained motion estimation," in *Proc. Visual Communication Image Processing (VCIP'94)*, A. K. Katsaggelos, Ed., Sept. 1994, pp. 1026–1034.
[20] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proc. ICASSP-93*, Minneapolis, MN, Apr. 1993, pp. 437–440.
[21] C. Ayeung, J. Kosmach, M. Orchard, and T. Kalafatis, "Overlapped block motion compensation," in *SPIE Conf. Visual Commun. Image Proc.*, Nov. 1992, pp. 561–571.
[22] H. Watanabe and S. Singhal, "Windowed motion compensation," in *Proc. SPIE VCIP-91*, Nov. 1991, pp. 582–589.
[23] S. Nogaki and M. Otha, "An overlapped block motion compensation for high quality motion picture coding," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1992, pp. 184–187.
[24] M. T. Orchard and G. J. Sullivan, "Overlapped block motion compensation: An estimation-theoretic approach," *IEEE Trans. Image Processing*, vol. 3, pp. 693–699, Sept. 1994.
[25] S.-W. Wu and A. Gersho, "Joint estimation of forward and backward motion vectors for interpolative prediction of video," *IEEE Trans. Image Processing*, vol. 3, pp. 684–687, Sept. 1994.
[26] M. Kaneko, Y. Hatori, and A. Kloike, "Improvements of transform coding algorithm for motion-compensated interframe prediction errors—DCT/SQ coding," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 1068–1078, Aug. 1987.
[27] M. Gilge, "A high quality videophone coder using hierarchical motion estimation and structure coding of the prediction error," in *Proc. SPIE Conf. Visual Commun. Image Proc.'88*, Cambridge, MA, Nov. 1998, pp. 864–874.
[28] P. Strobach, "Tree-structured scene-adaptive coder," *IEEE Trans. Commun.*, vol. 38, pp. 477–486, Apr. 1990.
[29] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
[30] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[31] R. Buschmann, "Efficiency of displacement estimation techniques," *Signal Process.: Image Commun.*, vol. 10, pp. 43–61, 1997.
[32] M. Flierl, T. Wiegand, and B. Girod, "A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction," in *Proc. Data Compression Conf.*, Snowbird, UT, Apr. 1998.

Bernd Girod (M'80–SM'97–F'98) received the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1980 and the Dr.(Hon.) degree from the University of Hannover, Hannover, Germany, in 1987.

Until 1987, he was a Member of the Research Staff at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, University of Hannover, working on moving image coding, human visual perception, and information theory. In 1988, he joined the Massachusetts Institute of Technology, Cambridge, first as a Visiting Scientist with the Research Laboratory of Electronics, then as an Assistant Professor of media technology at the Media Laboratory. From 1990 to 1993, he was Professor of Computer Graphics and Technical Director of the Academy of Media Arts, Cologne, Germany. He was a Visiting Adjunct Professor with the Digital Signal Processing Group at Georgia Institute of Technology in 1993. From 1993 to 1999, he was Chaired Professor of Electrical Engineering/Telecommunications at the University of Erlangen-Nuremberg, Germany, and the Head of the Telecommunications Institute I. He has served as the Chairman of the Electrical Engineering Department from 1995 to 1997, and as Director of the Center of Excellence "3-D Image Analysis and Synthesis" since 1995. He was a Visiting Professor with the Information Systems Laboratory, Stanford University, Stanford, CA, during the 1997–98 academic year. Since 2000, he has been a Professor of Electrical Engineering and Hoover Faculty Scholar with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University. His research interests include image communication, video signal compression, human and machine vision, computer graphics and animation, as well as interactive media. For several years, he has served as a consultant to government agencies and companies, with special emphasis on start-up ventures. He was Founder and Chief Scientist of Vivo Software, Inc., Waltham, MA, from 1993 to 1998. He has been Chief Scientist of RealNetworks, Inc., Seattle, WA, since 1998, and a Board Member of 8 × 8, Inc., Santa Clara, CA, since 1996. He has authored or co-authored one major textbook and more than 150 book chapters, journal articles, and conference papers in his field, and he holds several international patents.

Dr. Girod was an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1991 to 1995 and has been Reviewing Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS since 1995. He was a Member of the the Editorial Board of *Visual Communication and Image Representation* from 1993 to 1996, and member of the editorial board of *Computer and Graphics* from 1992 to 1999. He is a member of the editorial boards of *EURASIP Signal Processing*, *EURASIP Signal Processing: Image Communication*, IEEE SIGNAL PROCESSING MAGAZINE, and the *ACM Mobile Computing and Communication Review*. He chaired the 1990 SPIE Conference on Sensing and Reconstruction of Three-Dimensional Objects and Scenes, Santa Clara, CA, and the German Multimedia Conferences, Munich, in 1993 and 1994. He served as Tutorial Chair of ICASSP-97 in Munich and General Chair of the 1998 IEEE Image and Multidimensional Signal Processing Workshop, Alpbach, Austria. He will be the Tutorial Chair of ICIP-2000, Vancouver, BC, Canada, and General Chair of the Visual Communication and Image Processing Conference, San Jose, CA, in 2001.

Dr. Girod was a member of the IEEE Image and Multidimensional Signal Processing Committee from 1989 to 1997.