A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction

Markus Flierl, Thomas Wiegand

Telecommunications Laboratory
University of Erlangen-Nuremberg, Erlangen, Germany
{flierl,wiegand}@lnt.de

Bernd Girod

Information Systems Laboratory
Stanford University, Stanford, CA
girod@ee.stanford.edu

ABSTRACT

Multi-hypothesis prediction extends motion compensation with one prediction signal to the linear superposition of several motion-compensated prediction signals. These motion-compensated prediction signals are referenced by motion vectors and picture reference parameters. This paper proposes a state-of-the-art video codec based on the ITU-T Recommendation H.263 that incorporates multi-hypothesis motion-compensated prediction. In contrast to B-Frames, reference pictures are always previously decoded pictures. It is demonstrated that two hypotheses are efficient for practical video compression algorithms. In addition, it is shown that multi-hypothesis motion-compensated prediction and variable block size prediction can be combined to improve the overall coding gain. The encoder utilizes rate-constrained coder control including rate-constrained multi-hypothesis motion estimation. The advanced 4-hypothesis codec improves coding efficiency up to 1.8 dB when compared to the advanced prediction codec with ten reference frames for the set of investigated test sequences.

Keywords: Visual Communications, Video Coding, Multiframe Prediction, Multi-Hypothesis Motion-Compensated Prediction, Motion Estimation

1. INTRODUCTION

Today's state-of-the-art video codecs incorporate motion-compensated prediction (MCP). Some of these codecs employ more than one MCP signal simultaneously. The term "multi-hypothesis motion compensation" has been coined for this approach.[1] A linear combination of multiple prediction hypotheses is formed to arrive at the actual prediction signal. Theoretical investigations[2] show that a linear combination of multiple prediction hypotheses can improve the performance of motion compensated prediction.

B-Frames, as they are employed in H.263[3] or MPEG, are an example of multi-hypothesis motion compensation where two motion-compensated signals are superimposed to reduce the bit-rate of a video codec. But the B-Frame concept has a significant drawback: prediction references pictures before and after the B-Picture. The associated delay may be unacceptable for interactive applications. To overcome this problem, we have proposed prediction algorithms[4,5] which superimpose multiple prediction signals selected from past frames only.

Selecting hypotheses from several past reference frames can be accomplished with the concept of long-term memory motion-compensated prediction[6] by extending each motion vector by a picture reference parameter. This additional reference parameter overcomes the restriction that a specific hypothesis has to be chosen from a certain reference frame. The additional reference parameter enables the multi-hypothesis motion estimator to find an efficient set of prediction signals employing any of the reference frames.

The presented video codec utilizes prediction hypotheses that are square blocks in the respective reference frames. It is known that motion-compensated prediction with blocks of variable size improves the efficiency of video compression algorithms.[7] But there is the open question whether the concept of variable block size can be successfully combined with the idea of multi-hypothesis motion-compensated prediction.

The outline of this paper is as follows: In Section 2, the video codec utilizing multi-hypothesis motion-compensated prediction is explained. In addition, the syntax extensions to the ITU-T Recommendation H.263[3] are outlined. In Section 3, the coder control with multi-hypothesis motion estimation is presented. Section 4 discusses a model for multi-hypothesis motion-compensated prediction[2] and incorporates optimal multi-hypothesis motion estimation. This analysis provides insight about the number of hypotheses that have to be combined for an efficient video compression algorithm. Section 5 presents experimental results and demonstrates the efficiency of multi-hypothesis motion-compensated prediction for video coding.

2. MULTI-HYPOTHESIS VIDEO CODEC

The presented multi-hypothesis video codec is based on a standard hybrid video codec as proposed in ITU-T Recommendation H.263.[3] Such a codec is depicted in Fig. 1. Motion-compensated prediction is utilized to generate a prediction signal $\hat{\mathbf{s}}$ from previously reconstructed frames $\mathbf{r}$ in order to reduce the bit-rate of the intra-frame encoder. For block-based MCP, one motion vector and one picture reference parameter which address the reference block on previously reconstructed frames are assigned to each block in the current frame.

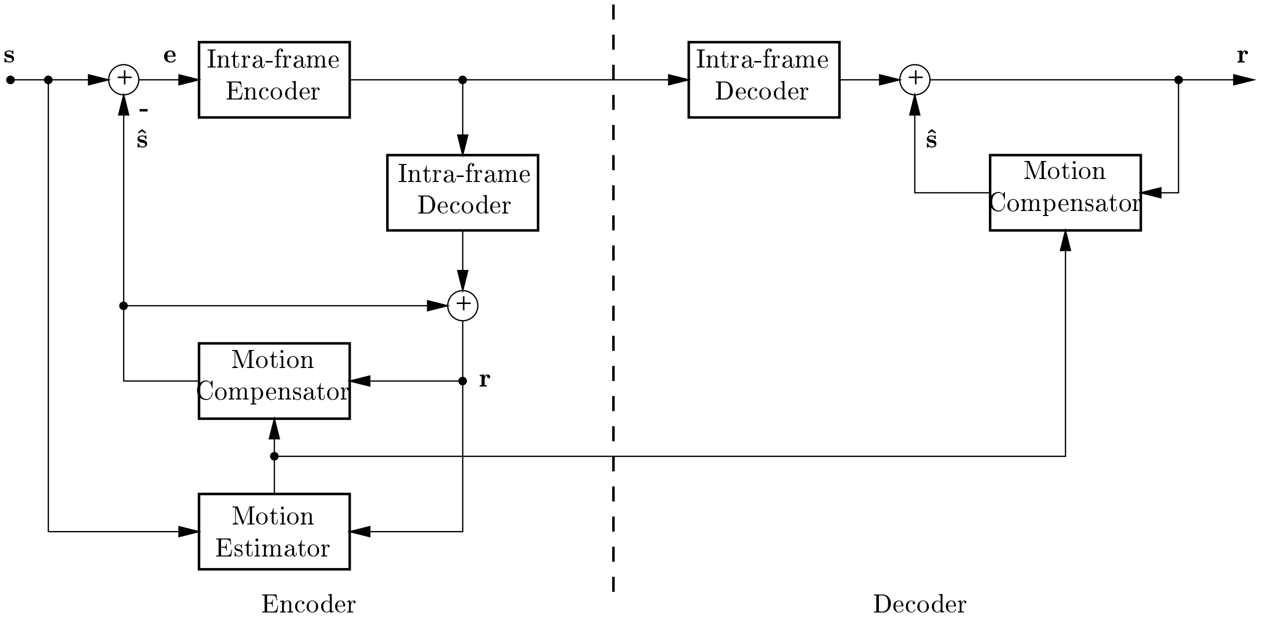

Figure 1. Standard video codec utilizing motion compensated prediction.

The multi-hypothesis video codec reduces the bit-rate of the intra-frame encoder even more by improving the prediction signal $\hat{\mathbf{s}}$. The improvement is achieved by linearly combining several motion-compensated prediction signals. For block-based multi-hypothesis MCP, several motion vectors and picture reference parameters which address more than one reference block in previous reconstructed frames are assigned to each block in the current frame. These multiple reference blocks are linearly combined to form the multi-hypothesis prediction signal.

2.1. Multi-Hypothesis Motion Compensation

Consider N motion-compensated signals $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_N$. We will refer to them as hypotheses. The multi-hypothesis prediction signal $\hat{\mathbf{s}}$ is the linear superposition of these N hypotheses. In general, each hypothesis can be weighted by a constant coefficient. Previous work on design of block-based multi-hypothesis motion-compensated predictors[4] suggests that simply averaging the hypotheses is efficient, i.e.,

$$\hat{\mathbf{s}} = \frac{1}{N}\sum_{\nu=1}^{N}\mathbf{c}_{\nu}. \tag{1}$$

Fig. 2 shows three hypotheses from previous decoded frames which are linearly combined to form the multi-hypothesis prediction signal for the current frame. Please note that a hypothesis can be chosen from any reference frame. Therefore, each hypothesis has to be assigned an individual picture reference parameter.

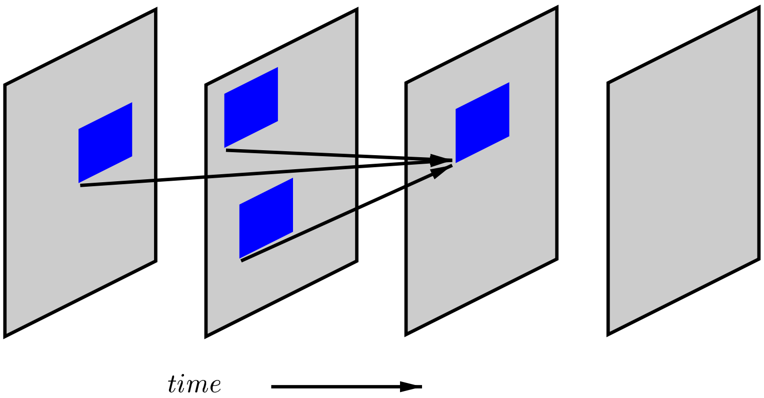

Figure 2. Multi-hypothesis motion-compensated prediction with three hypotheses. Three blocks of previous decoded frames are linearly combined to form a prediction signal for the current frame.

The proposed scheme differs from the concept of B-Frame prediction in three significant ways: First, all reference frames are past frames. No reference is made to a "future" frame, as with B-frames, and hence no extra delay is incurred. Second, hypotheses are not restricted to particular reference frames. This enables the encoder to find an efficient set of prediction signals. Third, it is possible to combine more than two motion-compensated signals. These three advantages of multi-hypothesis motion compensation improve the coding efficiency compared to an H.263 codec.

2.2. Syntax Extensions

The syntax of H.263 is extended such that multi-hypothesis motion compensation is possible. On the macroblock level, two new modes, INTER2H and INTER4H, are added which allow two or four hypotheses per macroblock, respectively. These modes are similar to the INTER mode of H.263. The INTER2H mode additionally includes an extra motion vector and frame reference parameter for the second hypothesis. The INTER4H mode incorporates three extra motion vectors and frame reference parameters. For advanced prediction, the INTER4V mode of H.263 is extended by a multi-hypothesis block pattern. This pattern indicates for each 8 × 8 block the number of motion vectors and frame reference parameters. This mode is called INTER4VMH. The multi-hypothesis block pattern has the advantage that the number of hypotheses can be indicated individually for each 8 × 8 block. This allows the important case that just one 8 × 8 block can be coded with more than one motion vector and frame reference parameter. The INTER4VMH mode includes the INTER4V mode when the multi-hypothesis block pattern indicates just one hypothesis for all 8 × 8 blocks.

3. CODER CONTROL

The coder control for the multi-hypothesis video codec utilizes rate-distortion optimization by Lagrangian methods. The target is to minimize the average Lagrangian costs for each frame, given the previous encoded frames,

$$J = D + \lambda R. \tag{2}$$

The average costs are constituted by the average distortion D and the weighted average bit-rate R. The weight, also called Lagrangian multiplier λ, is tied to the macroblock quantization parameter Q by the relationship[8]

$$\lambda = 0.85Q^2. \tag{3}$$

This generic optimization method provides the encoding strategy for the multi-hypothesis encoder: Minimizing the Lagrangian costs for each macroblock will minimize the average Lagrangian costs for each frame, given the previous encoded frames.

3.1. Multi-Hypothesis Motion Estimation

Multi-hypothesis motion compensation implies the estimation of multiple motion vectors and picture reference parameters. Best prediction performance is obtained when N motion vectors and picture reference parameters are jointly estimated. This joint estimation would be computationally very demanding. Complexity can be reduced by an iterative algorithm which improves conditional optimal solutions step by step.[4] The *Hypothesis Selection Algorithm* (HSA) as depicted in Fig. 3 is such an iterative algorithm. The HSA guarantees a local minimum for the instantaneous Lagrangian costs for each block in the current frame[4] and therefore performs rate-constrained multi-hypothesis motion estimation.

```
0: Assuming $N$ hypotheses $(\mathbf{c}_1, \ldots, \mathbf{c}_N)$, the rate-distortion cost function

       $j(\mathbf{c}_1, \ldots, \mathbf{c}_N) = \left\| \mathbf{s} - \frac{1}{N} \sum_{\nu=1}^{N} \mathbf{c}_\nu \right\|_2^2 + \lambda \sum_{\nu=1}^{N} r(\mathbf{c}_\nu)$

   is subject to minimization for each original block $\mathbf{s}$, given the Lagrange multiplier $\lambda$. Set $i := 0$ and guess $N$ initial hypotheses $(\mathbf{c}_1^{(0)}, \ldots, \mathbf{c}_N^{(0)})$.

1: Starting with the first and ending with the $N$-th hypothesis:

    a: Select the $\mu$-th hypothesis. All others are held constant.

    b: Minimize the rate-distortion cost function by full search for hypothesis $\mathbf{c}_\mu^{(i+1)}$

       $\min_{\mathbf{c}_\mu^{(i+1)}} j(\mathbf{c}_1^{(i+1)}, \ldots, \mathbf{c}_{\mu-1}^{(i+1)}, \mathbf{c}_\mu^{(i+1)}, \mathbf{c}_{\mu+1}^{(i)}, \ldots, \mathbf{c}_N^{(i)})$

2: As long as the rate-distortion cost function decreases, set $i := i + 1$ and continue with step 1.
```

Figure 3. The *Hypothesis Selection Algorithm* provides a locally optimal rate-constrained solution to the joint motion estimation problem.

3.2. Rate-Constrained Mode Decision

The new multi-hypothesis modes include both multi-hypothesis prediction and prediction error encoding. The Lagrangian costs of these new multi-hypothesis modes have to be evaluated in order to select the most efficient mode for each macroblock. The distortion of the reconstructed macroblock is determined by the summed squared error. The macroblock bit-rate includes also the rate of all motion vectors and picture reference parameters. This allows the best trade-off between multi-hypothesis MCP rate and prediction error rate.[9] Multi-hypothesis MCP improves the prediction signal by spending more bits for the side-information associated with the motion-compensating predictor. But the encoding of the prediction error and its associated bit-rate also determines the quality of the reconstructed block.

A joint optimization of multi-hypothesis motion estimation and prediction error encoding is far too demanding. But multi-hypothesis motion estimation independent of prediction error encoding is an efficient and practical solution. This solution is efficient if rate-constrained multi-hypothesis motion estimation (RC MH ME), as explained before, is applied.

For example, the encoding strategies for the INTER and INTER2H modes are as follows: Testing the INTER mode, the encoder performs successively rate-constrained motion estimation (RC ME) for integer-pel positions and rate-constrained half-pel refinement. RC ME incorporates the prediction error of the video signal as well as the bit-rate for the motion vector and picture reference parameter. Testing the INTER2H mode, the encoder performs rate-constrained multi-hypothesis motion estimation. RC MH ME incorporates the multi-hypothesis prediction error of the video signal as well as the bit-rate for two motion vectors and picture reference parameters. RC MH ME is performed by the HSA which utilizes in each iteration step RC ME to determine a conditional rate-constrained motion estimate. Given the obtained motion vectors and picture reference parameters for the INTER and INTER2H modes, the resulting prediction errors are encoded to evaluate the mode costs. The encoding strategy for the INTER4H mode is similar. For the INTER4VMH mode, the number of hypotheses for each 8 × 8 block has to be determined after encoding its residual error.

4. EFFICIENT NUMBER OF HYPOTHESES

An efficient video compression algorithm should trade off between the complexity of the algorithm and the achievable gain. The analysis in this section shows that, first, the gain by multi-hypothesis MCP with averaged hypotheses is theoretically limited even if the number of hypotheses grows infinite large and, second, the gain for two jointly optimized hypotheses is close to the theoretical limit. In the following, we focus on the dependency between multi-hypothesis prediction performance and displacement error correlation. Previous work on the subject can be found in Ref. 2.

4.1. Power Spectral Model for Inaccurate Multi-Hypothesis Motion Compensation

Let $\mathbf{s}[l]$ and $\mathbf{c}_\mu[l]$ be scalar two-dimensional signals sampled on an orthogonal grid with horizontal and vertical spacing of 1. $l = (x, y)^T$ denotes the vector valued location of the sample. For the problem of multi-hypothesis motion compensation, we interpret $\mathbf{c}_\mu$ as the μ-th of N motion-compensated frames available for prediction, and $\mathbf{s}$ as the current frame to be predicted. We call $\mathbf{c}_\mu$ also the μ-th hypothesis.

Obviously, multi-hypothesis 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. To capture the limited accuracy of motion compensation, we associate a vector valued displacement error $\mathbf{\Delta}_\mu$ with the μ-th hypothesis $\mathbf{c}_\mu$. 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. For simplicity, we assume that all hypotheses are shifted versions of the current frame signal $\mathbf{s}$. The shift is determined by the vector valued displacement error $\mathbf{\Delta}_\mu$ of the μ-th hypotheses. For that, the ideal reconstruction of the band-limited signal $\mathbf{s}[l]$ is shifted by the continuous valued displacement error and re-sampled on the original orthogonal grid. Our translatory displacement model omits "noisy" signal components which are also included in Ref. 2.

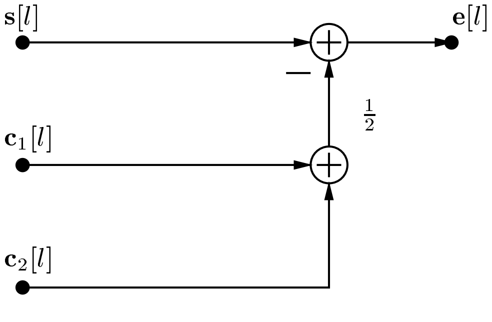

Figure 4. Multi-hypothesis motion-compensated prediction with two hypotheses. The current frame $\mathbf{s}[l]$ is predicted by averaging two hypotheses $\mathbf{c}_1[l]$ and $\mathbf{c}_2[l]$.

Fig. 4 depicts the predictor which averages two hypotheses $\mathbf{c}_1[l]$ and $\mathbf{c}_2[l]$ in order to predict the current frame $\mathbf{s}[l]$. In general, the prediction error for each pel at location l is the difference between the current frame signal and N averaged hypotheses

$$\mathbf{e}[l] = \mathbf{s}[l] - \frac{1}{N}\sum_{\mu=1}^{N} \mathbf{c}_\mu[l]. \tag{4}$$

Assume that $\mathbf{s}$ and $\mathbf{c}_\mu$ are generated by a jointly wide-sense stationary random process with the real-valued scalar two-dimensional power spectral density $\Phi_{\mathbf{ss}}(\omega)$ as well as the cross spectral densities $\Phi_{\mathbf{c}_\mu\mathbf{s}}(\omega)$ and $\Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega)$. Power spectra and cross spectra are defined according to

$$\Phi_{\mathbf{ab}}(\omega) = \mathcal{F}_* \left\{ E\left\{ \mathbf{a}[l_0 + l]\mathbf{b}^*[l_0] \right\} \right\} \tag{5}$$

where $\mathbf{a}$ and $\mathbf{b}$ are complex signals, $\mathbf{b}^*$ is the complex conjugate of $\mathbf{b}$, and $l \in \Pi$ are the sampling locations. $\phi_{\mathbf{ab}}[l] = E\left\{ \mathbf{a}[l_0 + l]\mathbf{b}^*[l_0] \right\}$ is the scalar space-discrete cross correlation function between the signals $\mathbf{a}$ and $\mathbf{b}$ which (for wide-sense stationary random processes) does not depend on l_0 but only on the relative two-dimensional shift l. Finally, $\mathcal{F}_*\{\cdot\}$ is the 2D band-limited discrete-space Fourier transform

$$\mathcal{F}_* \left\{ \phi_{\mathbf{ab}}[l] \right\} = \sum_{l \in \Pi} \phi_{\mathbf{ab}}[l] e^{-j\omega^T l} \quad \forall \quad \omega \in \left] -\pi, \pi \right] \times \left] -\pi, \pi \right] \tag{6}$$

where $\omega^T = (\omega_x, \omega_y)$ is the transpose of the vector valued frequency ω.

The power spectral density of the prediction error in (4) is given by the power spectrum of the current frame and the cross spectra of the hypotheses

$$\Phi_{\mathbf{ee}}(\omega) = \Phi_{\mathbf{ss}}(\omega) - \frac{2}{N} \sum_{\mu=1}^{N} \Re\left\{ \Phi_{\mathbf{c}_\mu\mathbf{s}}(\omega) \right\} + \frac{1}{N^2} \sum_{\mu=1}^{N} \sum_{\nu=1}^{N} \Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega), \tag{7}$$

where $\Re\{\cdot\}$ denotes the real component of the in general complex valued cross spectral densities $\Phi_{\mathbf{c}_\mu\mathbf{s}}(\omega)$. We adopt the expressions for the cross spectra from Ref. 2, where the displacement errors $\mathbf{\Delta}_\mu$ are interpreted as random variables which are statistically independent from $\mathbf{s}$:

$$\begin{aligned} \Phi_{\mathbf{c}_\mu\mathbf{s}}(\omega) &= \Phi_{\mathbf{ss}}(\omega) E\left\{ e^{-j\omega^T \mathbf{\Delta}_\mu} \right\} && (8) \\ \Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega) &= \Phi_{\mathbf{ss}}(\omega) E\left\{ e^{-j\omega^T (\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)} \right\} && (9) \end{aligned}$$

4.2. Assumptions about the Displacement Error PDF

For $\mathbf{\Delta}_\mu$, a 2-D stationary normal distribution with variance $\sigma^2_{\mathbf{\Delta}}$ and zero mean is assumed where x- and y-components are statistically independent.[2] The displacement error variance is the same for all N hypotheses. This is reasonable because all hypotheses are compensated with the same accuracy. Further, the pairs $(\mathbf{\Delta}_\mu, \mathbf{\Delta}_\nu)$ are assumed to be jointly Gaussian random variables. As there is no preference among the N hypotheses, the correlation coefficient $\rho_{\mathbf{\Delta}}$ between two displacement error components $\mathbf{\Delta}_{x\mu}$ and $\mathbf{\Delta}_{x\nu}$ is the same for all pairs of hypotheses. The above assumptions are summarized by the covariance matrix of a displacement error component.

$$C_{\mathbf{\Delta}_x\mathbf{\Delta}_x} = \sigma^2_{\mathbf{\Delta}} \begin{pmatrix} 1 & \rho_{\mathbf{\Delta}} & \cdots & \rho_{\mathbf{\Delta}} \\ \rho_{\mathbf{\Delta}} & 1 & \cdots & \rho_{\mathbf{\Delta}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\mathbf{\Delta}} & \rho_{\mathbf{\Delta}} & \cdots & 1 \end{pmatrix}. \tag{10}$$

It is well known, that the covariance matrix is nonnegative definite.[10] As a consequence, the correlation coefficient $\rho_{\mathbf{\Delta}}$ has the limited range

$$\frac{1}{1-N} \le \rho_{\mathbf{\Delta}} \le 1 \quad \text{for} \quad N = 2, 3, 4, \ldots, \tag{11}$$

which is dependent on the number of hypotheses N. In contrast to the work in Ref. 2, we do not assume that the displacement errors $\mathbf{\Delta}_\mu$ and $\mathbf{\Delta}_\nu$ are mutually independent for $\mu \neq \nu$.

This assumptions allow us to express the expected values in (8) and (9) in terms of the 2-D Fourier transform P of the continuous 2-D probability density function of the displacement error $\mathbf{\Delta}_\mu$.

$$E\left\{ e^{-j\omega^T \mathbf{\Delta}_\mu} \right\} = \int_{\mathcal{R}^2} p_{\mathbf{\Delta}_\mu}(\Delta) e^{-j\omega^T \Delta} d\Delta = e^{-\frac{1}{2}\omega^T \omega \sigma^2_{\mathbf{\Delta}}} = P(\omega, \sigma^2_{\mathbf{\Delta}}) \tag{12}$$

The expected value in (9) contains differences of two Gaussian random variables. It is known that the difference of two Gaussian random variables is also Gaussian and the variance is given by $\sigma^2 = 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})$. Therefore, we obtain for the expected value in (9)

$$E\left\{e^{-j\omega^T(\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)}\right\} = P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})\right) \quad \text{for} \quad \mu \neq \nu. \tag{13}$$

For $\mu = \nu$, the expected value in (9) is equal to one. With that, we obtain for the power spectrum of the prediction error in (7)

$$\frac{\Phi_{\mathbf{ee}}(\omega)}{\Phi_{\mathbf{ss}}(\omega)} = \frac{N+1}{N} - 2P(\omega, \sigma^2_{\mathbf{\Delta}}) + \frac{N-1}{N} P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})\right). \tag{14}$$

Setting $\rho_{\mathbf{\Delta}} = 0$ provides a result which is already reported in Ref. 2. Like in Ref. 2, we will assume a power spectrum $\Phi_{\mathbf{ss}}$ that corresponds to an exponentially decaying isotropic autocorrelation function with a correlation coefficient $\rho_{\mathbf{s}}$.

4.3. Optimal Multi-Hypothesis Motion Estimation

The displacement error correlation coefficient influences the performance of MH motion compensation. An optimal MH motion estimation algorithm will select sets of hypotheses that optimize the performance of MH motion compensation. In the following, we focus on the relationship between the prediction error variance

$$\sigma^2_{\mathbf{e}} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \Phi_{\mathbf{ee}}(\omega) d\omega \tag{15}$$

and the displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$. The prediction error variance is a useful measure because it is related to the minimum achievable transmission bit-rate.[2] Fig. 5 depicts the functional dependency of the normalized prediction error variance from the displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$ within the range (11). The dependency is plotted for $N = 2, 4, 8$, and ∞. The left plot depicts the dependency for very accurate motion compensation ($\sigma^2_{\mathbf{\Delta}} = 1/3072$), the right plot for very inaccurate motion compensation ($\sigma^2_{\mathbf{\Delta}} = 4/3$). The correlation coefficient of the frame signal $\rho_{\mathbf{s}} = 0.93$.[2] Reference is the prediction error variance of the single-hypothesis predictor $\sigma^2_{\mathbf{e},1}$. We observe in both plots that a decreasing correlation coefficient lowers the prediction error variance. (14) implies that this observation holds for any displacement error variance. Fig. 5 shows also that identical displacement errors ($\rho_{\mathbf{\Delta}} = 1$), and consequently, identical hypotheses will not improve single-hypothesis motion compensation.

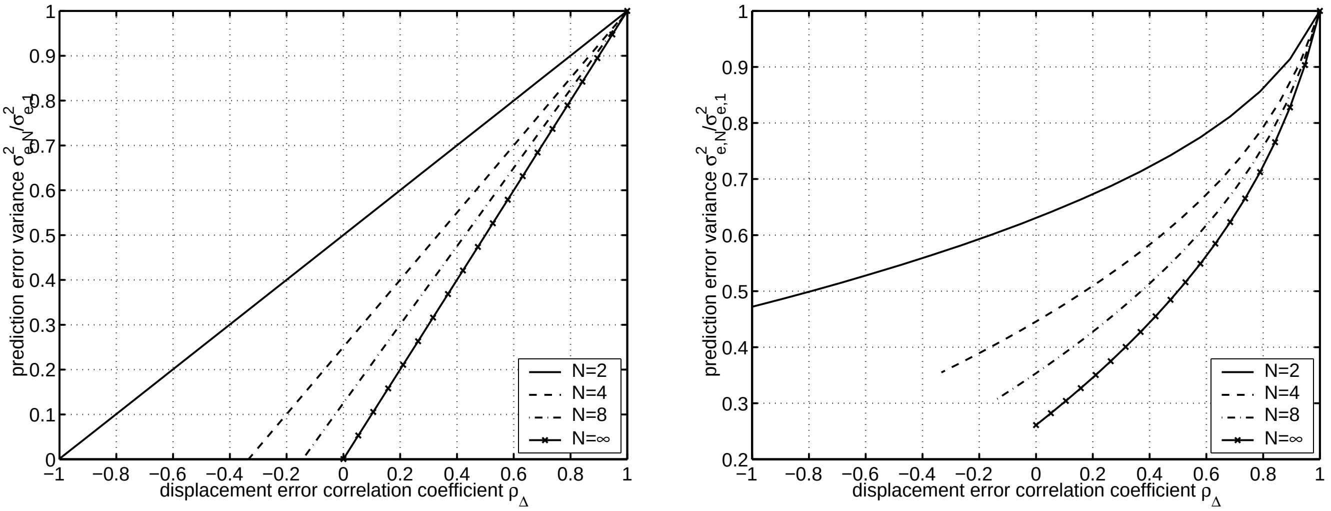

Figure 5. Normalized prediction error variance for multi-hypothesis MCP over the displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$. Reference is the single-hypothesis predictor. The hypotheses are averaged and no residual noise is assumed. The left plot depicts the dependency for $\sigma^2_{\mathbf{\Delta}} = 1/3072$, the right for $\sigma^2_{\mathbf{\Delta}} = 4/3$.

An optimal multi-hypothesis motion estimator minimizes for $\lambda = 0$ not only the summed squared error but also its expected value.[4] If a stationary error signal is assumed, this optimal estimator minimizes the prediction error variance. That is, an optimal multi-hypothesis motion estimator minimizes the prediction error variance by minimizing the displacement error correlation coefficient. Its minimum is given by the lower bound of the range (11).

$$\rho_{\boldsymbol{\Delta}} = \frac{1}{1-N} \quad \text{for} \quad N = 2, 3, 4, \ldots \tag{16}$$

This insight implies an interesting result for the case $N = 2$: Two jointly optimized hypotheses show the property that their displacement errors are maximally negatively correlated. The combination of two complementary hypotheses is more efficient than two independent hypotheses

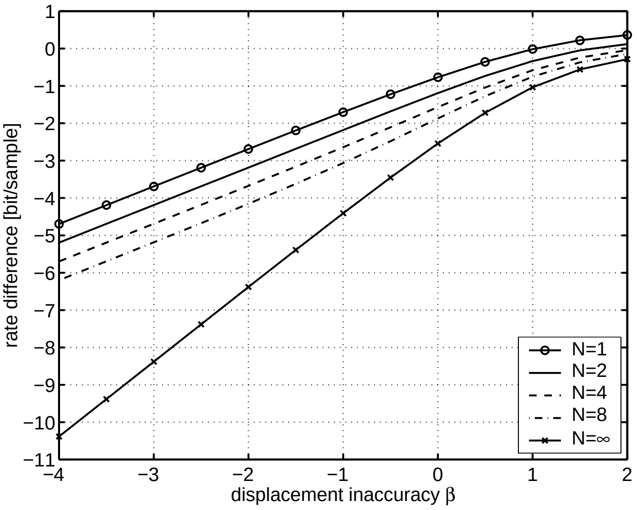

Figure 6. Rate difference for multi-hypothesis MCP over the displacement inaccuracy β for statistically independent displacement errors. The hypotheses are averaged and no residual noise is assumed.

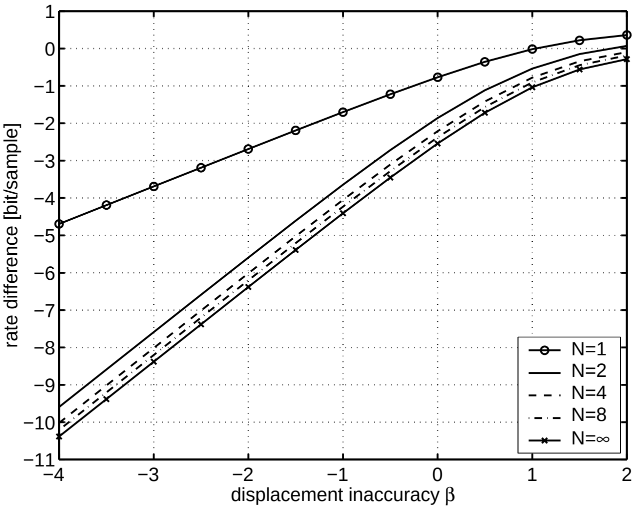

Figure 7. Rate difference for multi-hypothesis MCP over the displacement inaccuracy β for optimized displacement error correlation. The hypotheses are averaged and no residual noise is assumed.

Fig. 6 depicts the rate difference for multi-hypothesis motion-compensated prediction over the displacement inaccuracy β for statistically independent displacement errors according to Ref. 2. The rate difference[2]

$$\Delta R = \frac{1}{8\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \log_2 \left(\frac{\Phi_{\mathbf{ee}}(\omega)}{\Phi_{\mathbf{ss}}(\omega)} \right) d\omega \tag{17}$$

represents the maximum bit-rate reduction (in bit/sample) possible by optimum encoding of the prediction error $\mathbf{e}$, compared to optimum intra-frame encoding of the signal $\mathbf{s}$ for Gaussian wide-sense stationary signals for the same mean squared reconstruction error. A negative ΔR corresponds to a reduced bit-rate compared to optimum intra-frame coding. The maximum bit-rate reduction can be fully realized at high bit-rates, while for low bit-rates the actual gain is smaller.[2] The horizontal axis in Fig. 6 is calibrated by $\beta = \log_2(\sqrt{12}\sigma_{\boldsymbol{\Delta}})$. It is assumed that the displacement error is entirely due to rounding and is uniformly distributed in the interval $[-2^{\beta-1}, 2^{\beta-1}] \times [-2^{\beta-1}, 2^{\beta-1}]$, where $\beta = 0$ for integer-pel accuracy, $\beta = -1$ for half-pel accuracy, $\beta = -2$ for quarter-pel accuracy, etc.[2] The displacement error variance is

$$\sigma_{\boldsymbol{\Delta}}^2 = \frac{2^{2\beta}}{12}. \tag{18}$$

We observe in Fig. 6 that doubling the number of hypotheses decreases the bit-rate up to 0.5 bit per sample and the slope reaches up to 1 bit per sample and inaccuracy step. The case $N \to \infty$ achieves a slope up to 2 bit per sample and inaccuracy step. This can also be observed in (14) for $N \to \infty$ when we apply a Taylor series expansion of second order for the function P.

$$\frac{\Phi_{\mathbf{ee}}(\omega)}{\Phi_{\mathbf{ss}}(\omega)} \approx \sigma_{\boldsymbol{\Delta}}^4 \frac{1}{4} \left(\omega^T \omega\right)^2 \quad \text{for} \quad \sigma_{\boldsymbol{\Delta}}^2 \to 0, N \to \infty, \rho_{\Delta} = 0 \tag{19}$$

Inserting this result in (17) supports the observation in Fig. 6.

$$\Delta R \approx 2\beta + const. \quad \text{for} \quad \sigma^2_{\mathbf{\Delta}} \to 0, N \to \infty, \rho_\Delta = 0 \tag{20}$$

Fig. 7 depicts the rate difference for multi-hypothesis motion-compensated prediction over the displacement inaccuracy β for optimized displacement error correlation according to (16). We observe for accurate motion compensation that the slope of the rate difference of 2 bit per sample and inaccuracy step is already reached for $N = 2$. For increasing number of hypotheses the rate difference converges to the case $N \to \infty$ at constant slope. This suggests that a practical video coding algorithm should utilize two jointly optimized hypotheses.

5. EXPERIMENTAL RESULTS

The multi-hypothesis codec is based on the ITU-T Recommendation H.263[3] with Annexes D, F, and U. Ten reference pictures are used with the sliding window buffering mode. In contrast to H.263, a joint entropy code for horizontal and vertical motion vector data as well as a entropy code for the picture reference parameter is used. The test sequences are coded at QCIF resolution and 10 fps. Each sequence has a length of ten seconds. For comparison purposes, the PSNR values of the luminance component are measured and plotted over the total bit-rate for quantizer values of 4, 5, 7, 10, 15, and 25. The data of the first intra-frame coded picture are excluded from the results.

5.1. Efficiency of the Multi-Hypothesis Modes

The coding efficiency of the INTER2H and INTER4H modes are evaluated by allowing three different macroblock modes: The baseline coder includes the INTRA and INTER modes, the 2-hypothesis coder extends the baseline coder by the INTER2H mode, and the 4-hypothesis coder allows also the INTER4H mode. Figs. 8 and 9 depict the performance of the multi-hypothesis modes for the test sequences *Foreman* and *Mobile & Calendar*, respectively. A gain of up to 1 dB for the sequence *Foreman* and 1.4 dB for the sequence *Mobile & Calendar* is achieved by the INTER2H mode. An additional INTER4H mode gains just 0.1 dB for the sequence *Foreman* and 0.3 dB for the sequence *Mobile & Calendar.* This results also support the finding in the previous section that two hypotheses provide the largest relative gain.

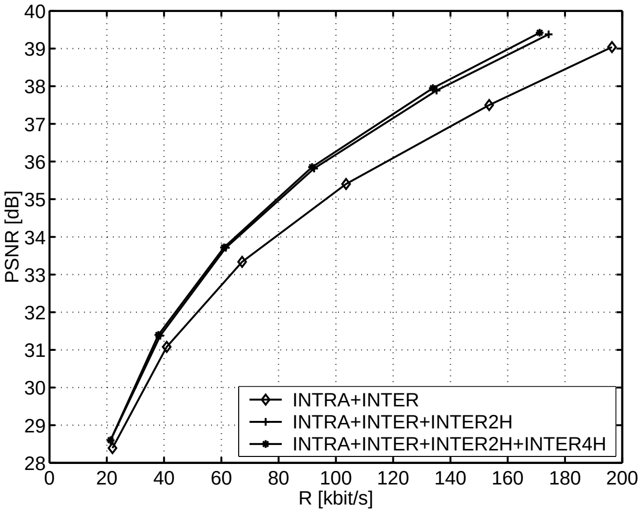

Figure 8. Average luminance PSNR over total rate for the sequence *Foreman* depicting the performance of the multi-hypothesis coding scheme.

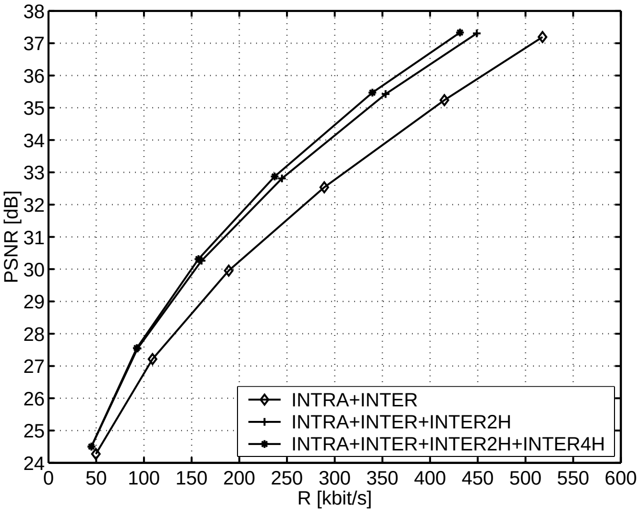

Figure 9. Average luminance PSNR over total rate for the sequence *Mobile & Calendar* depicting the performance of the multi-hypothesis coding scheme.

Figs. 10 and 11 compare the efficiency of the 2-hypothesis codec to the advanced prediction mode of H.263 without overlapped block motion compensation (OBMC) for the test sequences *Foreman* and *Mobile & Calendar.* The 2-hypothesis codec utilizes just two motion vectors and picture reference parameters compared to the advanced prediction mode with four motion vectors and picture references. But both codecs demonstrate comparable performance for these test sequences over the range of bit-rates considered.

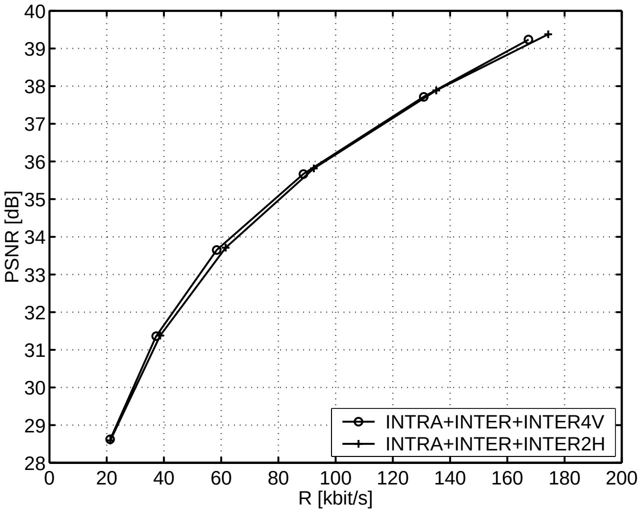

Figure 10. Average luminance PSNR over total rate for the sequence *Foreman*. The 2-hypothesis coder is compared to the advanced prediction coder.

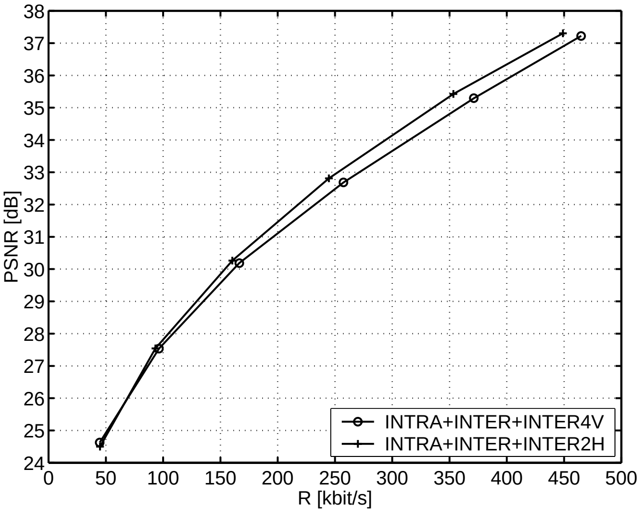

Figure 11. Average luminance PSNR over total rate for the sequence *Mobile & Calendar*. The 2-hypothesis coder is compared to the advanced prediction coder.

5.2. Multiple Hypotheses and Variable Block Size

Multi-hypothesis motion-compensated prediction is not only a competing scheme to motion-compensated prediction with variable block size. Multi-hypothesis prediction can also be combined with variable block sizes. Figs. 12 - 15 compare the advanced prediction mode of H.263 without OBMC to the advanced 2-hypothesis coder for the test sequences *Foreman*, *Mobile & Calendar*, *Sean*, and *Weather*. The advanced 2-hypothesis coder incorporates the modes INTRA, INTER, INTER2H, and INTER4VMH, where the multi-hypothesis block pattern allows for each 8×8 block one or two hypotheses.

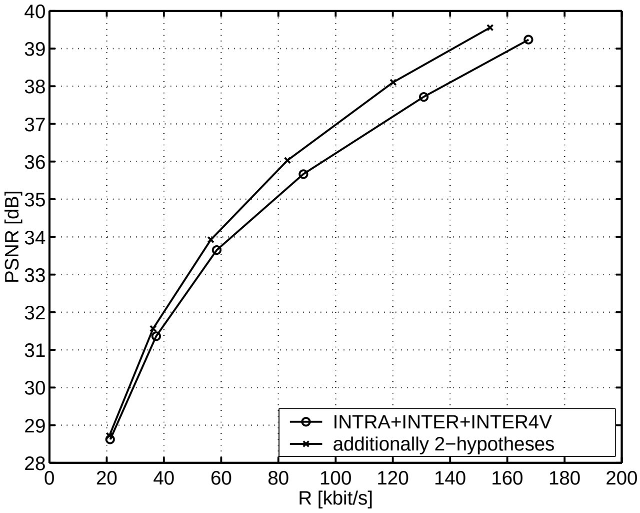

Figure 12. Average luminance PSNR over total rate for the sequence *Foreman* The advanced 2-hypothesis coder is compared to the advanced prediction coder.

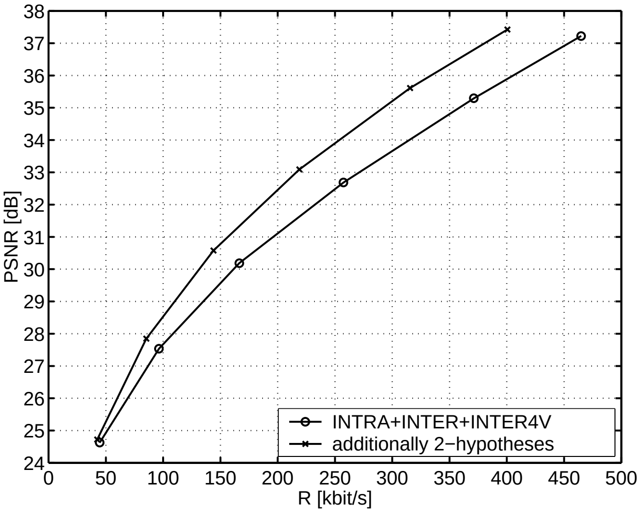

Figure 13. Average luminance PSNR over total rate for the sequence *Mobile & Calendar*. The advanced 2-hypothesis coder is compared to the advanced prediction coder.

It can be observed that the coding gains vary from 0.8 dB to 1.5 dB depending on the test sequence. For example, the coding gain for the sequence *Foreman* is up to 0.9 dB. (Fig. 12) The 2-hypothesis codec gains over the baseline codec for the same sequence up to 1 dB (Fig. 8). Similar observations can be made for the sequence *Mobile &*

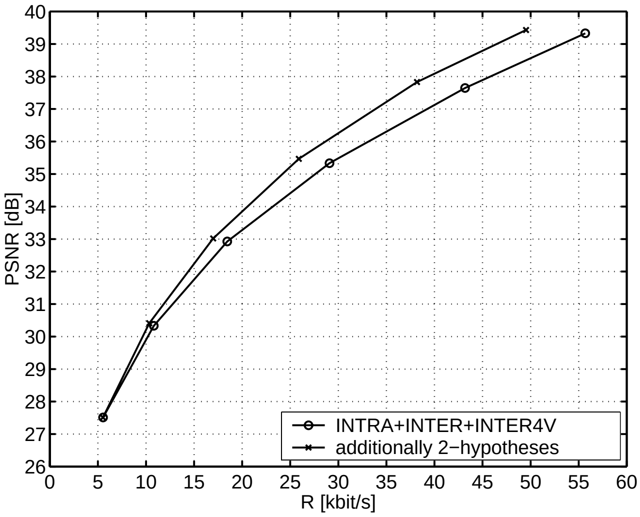

Figure 14. Average luminance PSNR over total rate for the sequence *Sean* The advanced 2-hypothesis coder is compared to the advanced prediction coder.

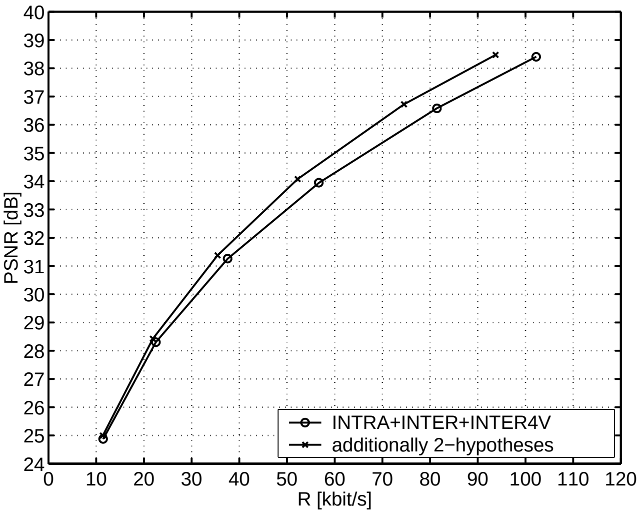

Figure 15. Average luminance PSNR over total rate for the sequence *Weather*. The advanced 2-hypothesis coder is compared to the advanced prediction coder.

Calendar with gains of 1.5 dB (Fig. 13) and 1.4 dB. (Fig. 9). It is concluded from this observation that the gains by multi-hypothesis prediction and variable block size prediction approximately add up.

Finally, it is noted that the results are obtained for multi-hypothesis motion-compensated prediction with ten reference pictures in sliding window buffering mode. But many video codecs employ just one reference picture, e.g. the TMN-10[11] coder, the test model of the H.263 standard. Figs. 16 and 17 show that the advanced prediction coder with ten reference pictures gains up to 0.9 dB for *Foreman* and 1.2 dB for *Mobile & Calendar* in comparison to the advanced prediction coder with just one reference picture. Comparing the advanced 2-hypothesis coder with ten reference pictures to the advanced prediction coder with one reference picture, coding gains up to 1.8 dB for *Foreman* and 2.7 dB for *Mobile & Calendar* can be achieved. Fig. 17 shows also that the advanced 4-hypothesis coder achieves the 3 dB gain for the sequence *Mobile & Calendar* .

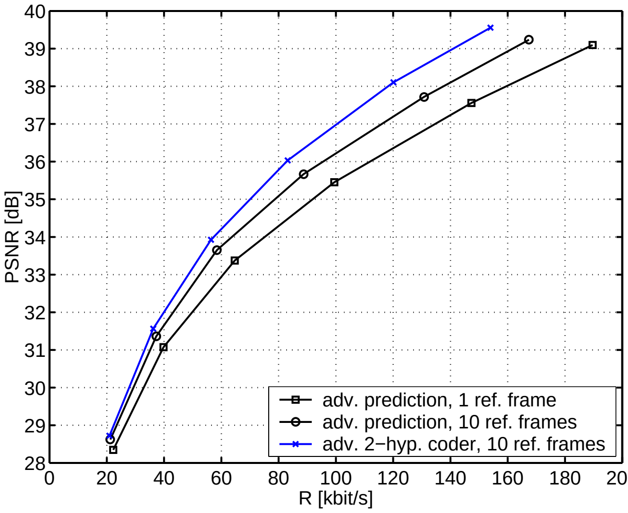

Figure 16. Average luminance PSNR over total rate for the sequence *Foreman* The advanced 2-hypothesis coder is compared to the advanced prediction coder with one reference picture.

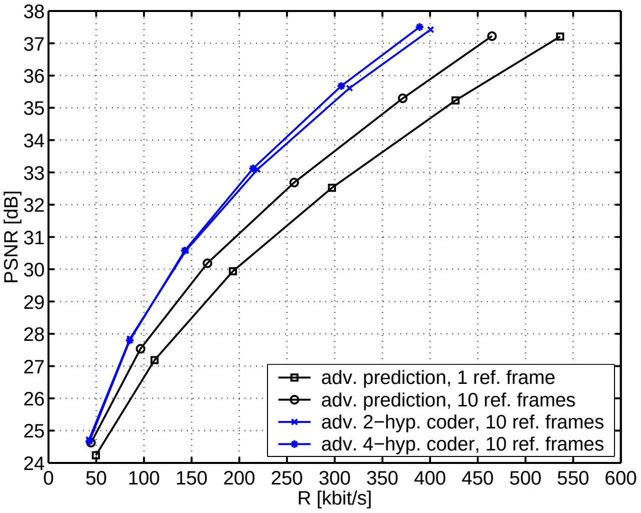

Figure 17. Average luminance PSNR over total rate for the sequence *Mobile & Calendar*. The advanced 4-hypothesis coder is compared to the advanced prediction coder with one reference picture.

6. CONCLUSIONS

Multi-hypothesis motion-compensated prediction[4] improves the coding efficiency of state-of-the-art video compression algorithms. The proposed scheme uses several motion vectors and picture reference parameters per block to address multiple prediction signals. These signals are averaged to form the prediction signal. Efficient encoding of video sequences is achieved by a rate-distortion optimized coder control with rate-constrained multi-hypothesis motion estimation. Rate-constrained multi-hypothesis motion estimation is performed by the *Hypothesis Selection Algorithm.* Theoretical investigations based on the previous analysis by Girod[2] suggest that a practical video coding algorithm should utilize two jointly optimized hypotheses. Experimental results verify this theoretical finding. Further, multi-hypothesis motion-compensated prediction and variable block size prediction can be combined with almost additive coding gains. The advanced 2-hypothesis coder with ten reference pictures achieves up to 2.7 dB coding gain compared to the advanced prediction coder with one reference picture for the sequence *Mobile & Calendar.*

REFERENCES

1. G. Sullivan, “Multi-Hypothesis Motion Compensation for Low Bit-Rate Video Coding,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Minneapolis*, vol. 5, pp. 437–440, Apr. 1993.
2. B. Girod, “Efficiency Analysis of Multi-Hypothesis Motion-Compensated Prediction for Video Coding.” IEEE Transactions on Image Processing, to appear Feb. 2000.
3. ITU-T, *Video Coding for Low Bitrate Communication: Recommendation H.263, Version 2*, 1998.
4. M. Flierl, T. Wiegand, and B. Girod, “A Locally Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-Compensated Prediction,” in *Proceedings of the Data Compression Conference, Snowbird, Utha*, pp. 239–248, Apr. 1998.
5. T. Wiegand, M. Flierl, and B. Girod, “Entropy-Constrained Linear Vector Prediction for Motion-Compensated Video Coding,” in *Proceedings of the International Symposium on Information Theory, Cambridge, MA*, p. 409, Aug. 1998.
6. T. Wiegand, X. Zhang, and B. Girod, “Long-Term Memory Motion-Compensated Prediction,” *IEEE Transactions on Circuits and Systems for Video Technology* **9**, pp. 70–84, Feb. 1999.
7. G. Sullivan and R. Baker, “Rate-Distortion Optimized Motion Compensation for Video Compression Using Fixed or Variable Size Blocks,” in *Proceedings of the IEEE Global Telecommunications Conference, Phoenix, AZ*, vol. 3, pp. 85–90, Dec. 1991.
8. G. Sullivan and T. Wiegand, “Rate-Distortion Optimization for Video Compression,” *IEEE Signal Processing Magazine* **15**, pp. 74–90, Nov. 1998.
9. B. Girod, “Rate-Constrained Motion Estimation,” in *SPIE Symposium on Visual Communications and Image Processing, Chicago*, pp. 1026–1034, Sept. 1994.
10. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.
11. ITU-T/SG16/Q15-D-65, *Video Codec Test Model, Near Term, Version 10 (TMN-10), Draft 1*, Apr. 1998. Download via anonymous ftp to: standard.pictel.com/video-site/9804_Tam/q15d65d1.doc.