Multihypothesis Motion Estimation for Video Coding

Markus Flierl†
Telecommunications Laboratory
University of Erlangen-Nuremberg
D-91058 Erlangen, Germany
flierl@LNT.de

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

Abstract

Multihypothesis motion-compensating predictors combine several motion-compensated signals to predict the current frame of a video signal. This paper applies the wide-sense stationary theory of multihypothesis motion compensation for hybrid video codecs to multihypothesis motion estimation. This allows us to study the influence of the displacement error correlation on the efficiency of multihypothesis motion compensation. Reducing the displacement error correlation between the hypotheses decreases the variance of the multihypothesis prediction error. We derive a property for the displacement error correlation coefficient for an optimal multihypothesis motion estimator in the mean squared error sense. We observe for the wide-sense stationary model that jointly optimal motion estimation improves the prediction performance and reduces the prediction error variance up to 12 dB per accuracy refinement step compared to 6 dB per accuracy refinement step for uncorrelated displacement errors. Consequently, the gain of multihypothesis motion-compensated prediction with jointly optimal motion estimation over motion-compensated prediction increases by improving the accuracy of each hypothesis. We also discuss the combination of hypotheses with additive noise and extend the predictor by the optimum Wiener filter.

1 Introduction

Efficient video compression algorithms employ more than one motion-compensated signal simultaneously to predict the current frame of a video signal. The term ”multihypothesis motion compensation” has been coined for this approach. Theoretical investigations in [1] show that a linear combination of multiple prediction hypotheses can improve the performance of motion-compensated prediction. It is reported in [2] that an optimal multihypothesis motion estimation algorithm selects hypotheses

†Markus Flierl currently visits the Information Systems Laboratory at Stanford University.

such that their displacement error correlation coefficient is maximally negative. In this paper, we will investigate this property in more detail.

A practical algorithm for rate-constrained multihypothesis motion estimation has been presented first in [3]. The iterative algorithm improves conditionally optimal solutions and provides a local optimum for the joint estimation problem. The experimental results demonstrate that the joint estimation of hypotheses is important for multihypothesis motion-compensated prediction.

The joint estimation of hypotheses is also important for a video coding scheme that utilizes multihypothesis motion-compensated prediction. It is shown in [4] that jointly estimated hypotheses improve not only the motion-compensated prediction signal but also the rate-distortion efficiency of a hybrid video codec.

The paper is organized as follows: Section 2 adopts the power spectral model for inaccurate multihypothesis motion compensation from [1] and extends it such that multihypothesis motion estimation can be investigated. Section 3 discusses optimal multihypothesis motion estimation and derives a property for the displacement error correlation coefficient. Section 4 investigates the influence of the frame signal model on the performance of multihypothesis motion-compensated prediction. Section 5 analyzes "noisy" hypothesis and presents the performance of the multihypothesis motion-compensated predictor with Wiener filter for jointly estimated hypothesis.

2 Power Spectral Model for Inaccurate Multihypothesis 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. The vector $l = (x, y)^T$ denotes the location of the sample. For the problem of multihypothesis 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, motion-compensated prediction should work best if we compensate the true displacement of the scene exactly for a 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 motion compensation and transmission. 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. For now, our translatory displacement model omits "noisy" signal components.

A multihypothesis motion-compensating predictor forms a prediction signal by averaging N hypotheses $\mathbf{c}_\mu[l]$ in order to predict the current frame signal $\mathbf{s}[l]$. The prediction error for each pel at location l is the difference between the current frame signal and N averaged hypotheses

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

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 $\Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega)$. The power spectral density of the prediction error in (1) is given by the power spectrum of the current frame and the cross spectra of the hypotheses

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

where $\Re\{\cdot\}$ denotes the real component of the, in general, complex valued cross spectral densities $\Phi_{\mathbf{c}_\mu\mathbf{s}}(\omega)$, and $\omega = (\omega_x, \omega_y)^T$ the vector-valued frequency. We adopt the expressions for the cross spectra from [1], where the displacement errors $\mathbf{\Delta}_\mu$ are interpreted as random variables which are statistically independent from $\mathbf{s}$:

$$\Phi_{\mathbf{c}_\mu\mathbf{s}}(\omega) = \Phi_{\mathbf{ss}}(\omega) E\left\{e^{-j\omega^T \mathbf{\Delta}_\mu}\right\} \tag{3}$$

$$\Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega) = \Phi_{\mathbf{ss}}(\omega) E\left\{e^{-j\omega^T (\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)}\right\} \tag{4}$$

Like in [1], we assume a power spectrum $\Phi_{\mathbf{ss}}$ that corresponds to an exponentially decaying isotropic autocorrelation function with a correlation coefficient $\rho_{\mathbf{s}}$.

For the μ-th displacement error $\mathbf{\Delta}_\mu$, a 2-D stationary normal distribution with variance $\sigma^2_{\mathbf{\Delta}}$ and zero mean is assumed where the x- and y-components are statistically independent. 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. The predictor design in [3] showed that there is no preference among the N hypotheses. Consequently, 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.

$$C_{\mathbf{\Delta}_x\mathbf{\Delta}_x} = \sigma^2_{\mathbf{\Delta}} \begin{pmatrix} 1 & \rho_{\mathbf{\Delta}} & \cdots & \rho_{\mathbf{\Delta}} \\ \rho_{\mathbf{\Delta}} & 1 & \cdots & \rho_{\mathbf{\Delta}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\mathbf{\Delta}} & \rho_{\mathbf{\Delta}} & \cdots & 1 \end{pmatrix}. \tag{5}$$

Since the covariance matrix is nonnegative definite [5], the correlation coefficient $\rho_{\mathbf{\Delta}}$ in (5) has the limited range

$$\frac{1}{1-N} \le \rho_{\mathbf{\Delta}} \le 1 \quad \text{for} \quad N = 2, 3, 4, \ldots, \tag{6}$$

which is dependent on the number of hypotheses N. In contrast to [1], we do not assume that the displacement errors $\mathbf{\Delta}_\mu$ and $\mathbf{\Delta}_\nu$ are mutually independent for $\mu \neq \nu$.

These assumptions allow us to express the expected values in (3) and (4) in terms of the 2-D Fourier transform P of the continuous 2-D probability density function of the displacement error $\mathbf{\Delta}_\mu$.

$$E\left\{e^{-j\omega^T \mathbf{\Delta}_\mu}\right\} := P(\omega, \sigma^2_{\mathbf{\Delta}}) = e^{-\frac{1}{2}\omega^T\omega\sigma^2_{\mathbf{\Delta}}} \tag{7}$$

The expected value in (4) contains differences of jointly Gaussian random variables. The difference of two jointly Gaussian random variables is also Gaussian. As the two random variables have equal variance $\sigma^2_{\mathbf{\Delta}}$, the variance of the difference signal is given by $\sigma^2 = 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})$. Therefore, we obtain for the expected value in (4)

$$E\left\{e^{-j\omega^T(\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)}\right\} = P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})\right) \quad \text{for} \quad \mu \neq \nu. \tag{8}$$

For $\mu = \nu$, the expected value in (4) is equal to one. With that, we obtain for the power spectrum of the prediction error in (2):

$$\frac{\Phi_{\mathbf{ee}}(\omega)}{\Phi_{\mathbf{ss}}(\omega)} = \frac{N+1}{N} - 2P(\omega, \sigma^2_{\mathbf{\Delta}}) + \frac{N-1}{N} P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})\right) \tag{9}$$

Setting $\rho_{\mathbf{\Delta}} = 0$ provides a result which is presented in [1].

3 Optimal Multihypothesis Motion Estimation

The previous section shows that the displacement error correlation coefficient influences the performance of multihypothesis motion compensation. In the following, we focus on the relationship between the prediction error variance

$$\sigma^2_{\mathbf{e}} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \Phi_{\mathbf{ee}}(\omega) d\omega \tag{10}$$

and the displacement error correlation coefficient.

Fig. 1 depicts the dependency of the normalized prediction error variance from the displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$ within the range (6). The dependency is plotted for $N = 2, 4, 8$, and ∞ for integer-pel accurate motion compensation ($\sigma^2_{\mathbf{\Delta}} = 1/12$). The correlation coefficient of the frame signal $\rho_{\mathbf{s}} = 0.93$ [1]. Reference is the prediction error variance of the single-hypothesis predictor $\sigma^2_{\mathbf{e},1}$. We observe that a decreasing correlation coefficient lowers the prediction error variance. (9) implies that this observation holds for any displacement error variance. Fig. 1 shows also that identical displacement errors ($\rho_{\mathbf{\Delta}} = 1$) will not reduce the prediction error variance compared to single-hypothesis motion compensation. This is reasonable when we consider identical hypotheses. They do not improve multihypothesis motion-compensation because they have identical displacement errors.

We will use the following definition: An *optimal multihypothesis motion estimator* is an algorithm that selects sets of hypotheses such that the performance of multihypothesis motion compensation is optimized.

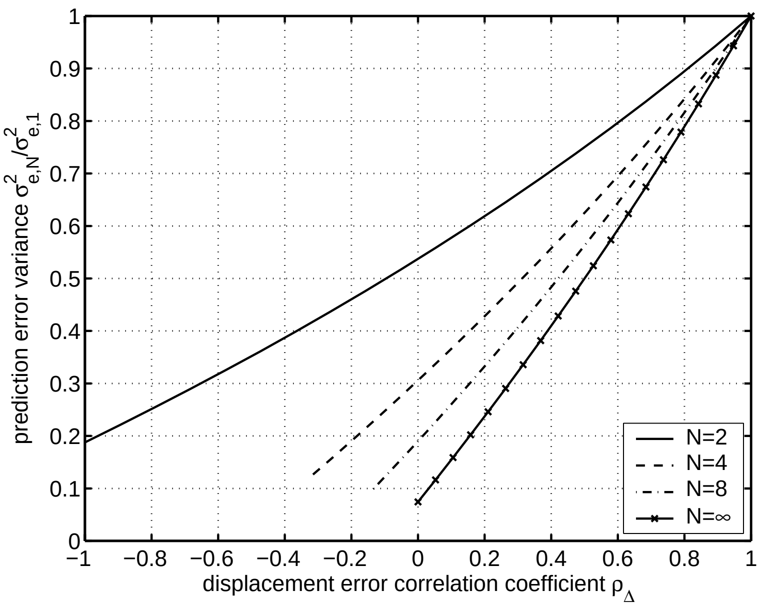

Figure 1: Normalized prediction error variance for multihypothesis MCP over the displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$. Reference is the single-hypothesis predictor. The hypotheses are averaged and no residual noise is assumed. The variance of the displacement error is set to $\sigma^2_{\mathbf{\Delta}} = 1/12$.

Assuming a mean squared error measure, the optimal multihypothesis motion estimator minimizes not only the summed squared error but also its expected value [3]. If a stationary error signal is assumed, this optimal estimator minimizes the prediction error variance. That is, an optimal multihypothesis 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 (6).

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

This insight implies an interesting result for the case $N = 2$: Two optimally jointly estimated 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.

The horizontal axis in Fig. 2 is calibrated by $\beta = \log_2(\sqrt{12}\sigma_{\mathbf{\Delta}})$. 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 [1]. The displacement error variance is

$$\sigma^2_{\mathbf{\Delta}} = \frac{2^{2\beta}}{12}. \tag{12}$$

Fig. 2 depicts the prediction error variance for multihypothesis motion-compensated prediction over the displacement inaccuracy β for both optimized displacement error correlation according to (11) and uncorrelated displacement errors. For

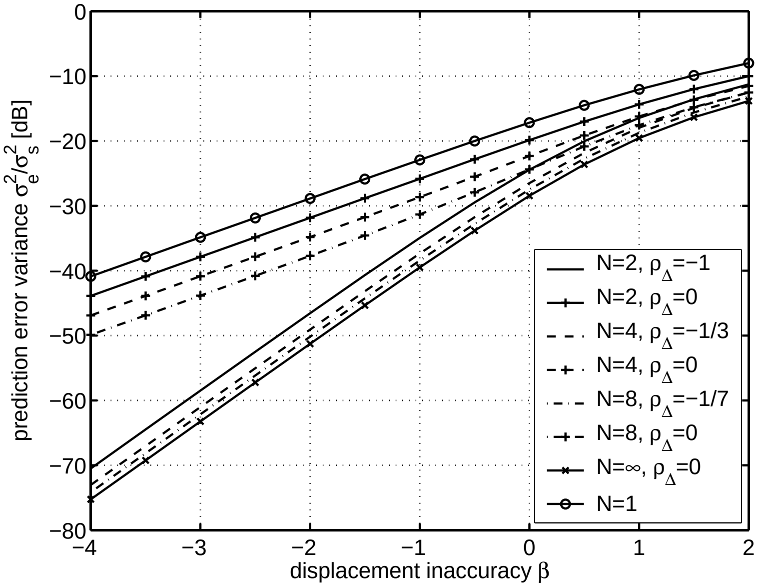

Figure 2: Prediction error variance for multihypothesis MCP over the displacement inaccuracy β for both optimized displacement error correlation and statistically independent displacement errors. The hypotheses are averaged and no residual noise is assumed.

uncorrelated displacement errors, we observe in Fig. 2 that doubling the number of hypotheses decreases the error variance up to 3 dB and the slope reaches up to 6 dB per inaccuracy step. The case $N \to \infty$ achieves a slope up to 12 dB per inaccuracy step. For optimized displacement error correlation, we observe for accurate motion compensation that the slope of 12 dB per inaccuracy step is already reached for $N = 2$. For increasing number of hypotheses the prediction error variance converges to the case $N \to \infty$ at constant slope. We obtain for the band-limited signal the following result: the gain of multihypothesis motion-compensated prediction with jointly optimal motion estimation over motion-compensated prediction increases by improving the accuracy of motion compensation for each hypothesis.

4 Influence of the Frame Signal Model

The model in Section 2 assumes an ideally sampled and reconstructed band-limited signal. The result in the last section originates from both the optimal joint estimation and the model assumptions. To show this, we alter the model assumptions for the frame signal.

For the comparison, we will assume a non-band-limited frame signal with the isotropic autocorrelation function

$$\phi_{\mathbf{ss}}(l) = \sigma_{\mathbf{s}}^2 e^{-a\sqrt{l_x^2+l_y^2}}. \tag{13}$$

(1) allows us to calculate directly the prediction error variance

$$\sigma_e^2 = \sigma_s^2 - \frac{2}{N}\sum_{\mu=1}^{N} E\left\{\phi_{\mathbf{ss}}(\mathbf{\Delta}_\mu)\right\} + \frac{1}{N^2}\sum_{\mu=1}^{N}\sum_{\nu=1}^{N} E\left\{\phi_{\mathbf{ss}}(\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)\right\}. \quad (14)$$

We still assume a multi-dimensional Gaussian probability density function for the displacement errors. This allows us to calculate a closed-form solution for the prediction error variance

$$\frac{\sigma_{\mathbf{e}}^2}{\sigma_{\mathbf{s}}^2} = \frac{N+1}{N} + \frac{N-1}{N} f(a\sqrt{2}\sigma_{\mathbf{\Delta}}\sqrt{1-\rho_{\mathbf{\Delta}}}) - 2f(a\sigma_{\mathbf{\Delta}}) \quad \text{for} \quad \frac{1}{1-N} \le \rho_{\mathbf{\Delta}} \le 1, \quad (15)$$

with the function

$$f(b) = 1 - \sqrt{\frac{\pi}{2}}\, b e^{\frac{1}{2}b^2} \operatorname{erfc}\left(\frac{b}{\sqrt{2}}\right), \quad (16)$$

containing the complementary error function $\operatorname{erfc}(\cdot)$. The optimal multihypothesis motion estimator minimizes the prediction error variance by minimizing the displacement error correlation coefficient. Its minimum is also given by the lower bound of (6).

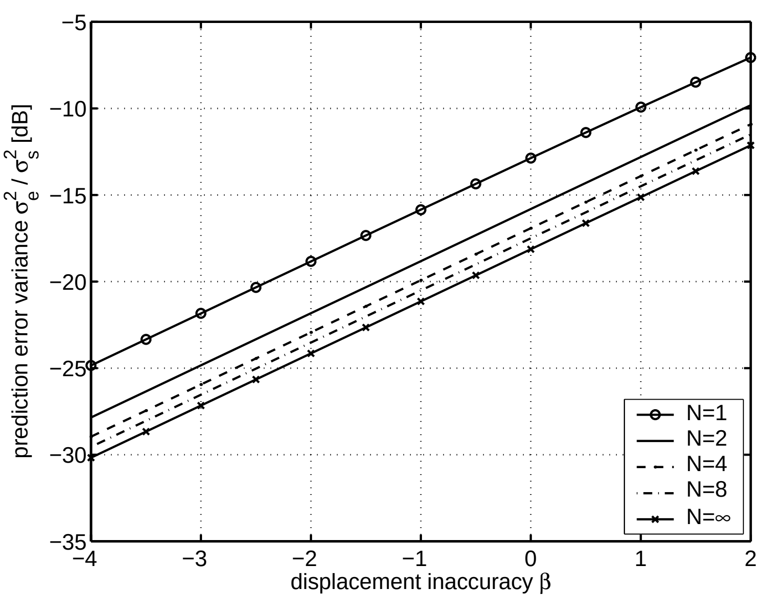

Figure 3: Performance of the optimal multihypothesis predictor for a non-band-limited frame signal. The hypotheses are averaged and no residual noise is assumed. $a = -\ln(0.93)$.

Fig. 3 depicts the performance of the optimal multihypothesis predictor for a non-band-limited frame signal. We observe that independent from the number of hypotheses the slope reaches up to 3 dB per inaccuracy step. The gain that we observed in Fig. 2 is also due to the band-limited character of the frame signal. Multihypothesis motion compensation with jointly optimal motion estimation is advantageous for band-limited signals.

5 Hypotheses with Additive Noise

In order to consider signal components that cannot be modeled by motion compensation, statistically independent noise $\mathbf{n}_\mu$ is added to each motion-compensated signal. In addition, an optimum Wiener filter is applied to all hypotheses. We also assume that the current frame originates from a "clean" video signal $\mathbf{v}$ with power spectral density $\Phi_{\mathbf{vv}}(\omega)$ [1]. For the case that the noise energy is the same for all hypotheses and the current frame, we obtain for the power spectral density of the prediction error

$$\frac{\Phi_{\mathbf{ee}}(\omega)}{\Phi_{\mathbf{ss}}(\omega)} = 1 - \frac{1}{1+\alpha(\omega)} \frac{P^2(\omega, \sigma^2_{\mathbf{\Delta}})}{P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})\right)} \frac{N}{N + \frac{1+\alpha(\omega)}{P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1-\rho_{\mathbf{\Delta}})\right)} - 1} \qquad (17)$$

with

$$\alpha(\omega) = \frac{\Phi_{\mathbf{n}_\mu \mathbf{n}_\mu}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} \quad \forall \mu. \qquad (18)$$

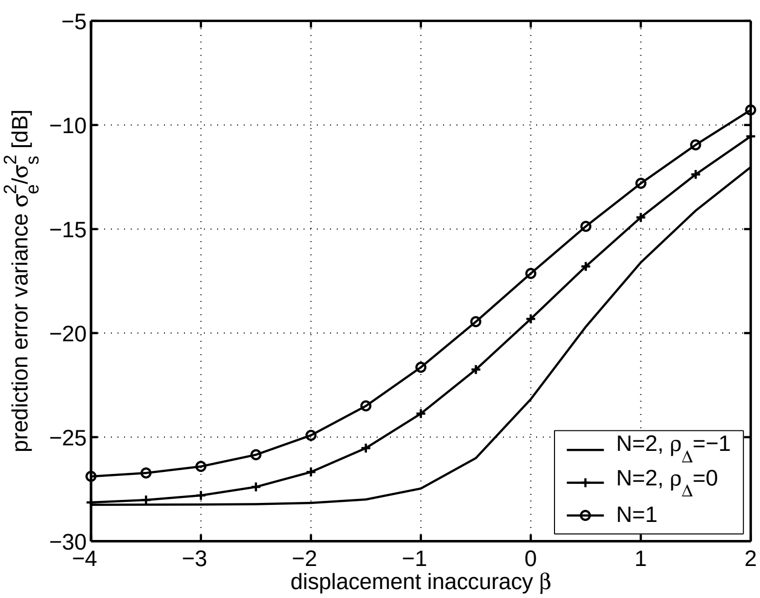

Figure 4: Prediction error variance for multihypothesis MCP over the displacement inaccuracy β for both optimized displacement error correlation and statistically independent displacement errors. Residual noise level RNL = -30 dB. In all cases, the optimum filter is applied.

Fig. 4 depicts the prediction error variance for multihypothesis MCP over the displacement inaccuracy β for both optimized displacement error correlation and statistically independent displacement errors. The residual noise level is chosen to be -30 dB. For half-pel accurate motion compensation and 2 hypotheses, we gain almost 4 dB in prediction error variance for optimized displacement error correlation over statistically independent displacement errors.

6 Conclusions

We investigate the performance of multihypothesis motion compensation for optimal multihypothesis motion estimation. Jointly estimated hypotheses show the property that their displacement errors are maximally negatively correlated. Hypotheses with negatively correlated displacement errors improve the performance of multihypothesis motion compensation.

For the investigated band-limited signal, the gain of multihypothesis motion-compensated prediction with jointly optimal motion estimation over motion-compensated prediction increases by improving the accuracy of motion compensation for each hypothesis.

For signals that are not band-limited, jointly optimal motion estimation also decreases prediction error variance but does not change the slope of the prediction error variance. In contrast to the ideally sampled and reconstructed band-limited signal, the multihypothesis gain over motion-compensated prediction is independent of the accuracy of motion compensation.

We also discuss the combination of hypotheses with additive noise and extend the predictor by the optimum Wiener filter. For "noisy" hypotheses, we obtain improved prediction performance for optimal multihypothesis motion estimation with the optimum Wiener filter.

7 Acknowledgment

The authors would like to thank Thomas Wiegand for helpful discussions and useful comments on the theory.

References

[1] B. Girod, "Efficiency Analysis of Multihypothesis Motion-Compensated Prediction for Video Coding", *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173–183, Feb. 2000.

[2] M. Flierl, T. Wiegand, and B. Girod, "A Video Codec Incorporating Block-Based Multi-Hypothesis Motion-Compensated Prediction", in *Proceedings of the SPIE Conference on Visual Communications and Image Processing*, Perth, Australia, June 2000, vol. 4067, pp. 238–249.

[3] 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, Utah, Apr. 1998, pp. 239–248.

[4] M. Flierl, T. Wiegand, and B. Girod, "Rate-Constrained Multi-Hypothesis Motion-Compensated Prediction for Video Coding", in *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, Canada, Sept. 2000, vol. III, pp. 150–153.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1991.