A Locally Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-Compensated Prediction

Markus Flierl, Thomas Wiegand, and Bernd Girod
Telecommunications Laboratory
University of Erlangen-Nuremberg
Cauerstr. 7, D-91058 Erlangen, Germany
{flierl|wiegand|girod}@nt.e-technik.uni-erlangen.de

Abstract

Multi-hypothesis motion-compensated prediction extends traditional motion-compensated prediction used in video coding schemes. Known algorithms for block-based multi-hypothesis motion-compensated prediction are, for example, overlapped block motion compensation (OBMC) and bidirectionally predicted frames (B-frames). This paper presents a generalization of these algorithms in a rate-distortion framework. All blocks which are available for prediction are called hypotheses. Further, we explicitly distinguish between the search space and the superposition of hypotheses. Hypotheses are selected from a search space and their spatio-temporal positions are transmitted by means of spatio-temporal displacement codewords. Constant predictor coefficients are used to combine linearly hypotheses of a multi-hypothesis. The presented design algorithm provides an estimation criterion for optimal multi-hypotheses, a rule for optimal displacement codes, and a condition for optimal predictor coefficients. Statistically dependent hypotheses of a multi-hypothesis are determined by an iterative algorithm. Experimental results show that increasing the number of hypotheses from 1 to 8 provides prediction gains up to 3 dB in prediction error.

1 Introduction

Motion-compensated coding schemes achieve data 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. Successful applications range from digital video broadcasting to low rate videophones. Several standards, such as ITU-T H.263, are based on this scheme.

Many codecs today employ more than one motion-compensated prediction signal simultaneously to predict the current frame. The term "multi-hypothesis motion compensation" has been coined for this approach. 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 or overlapped block motion compensation in the MPEG or H.263 coding schemes.

Performance bounds of multi-hypothesis MCP are investigated in [1] by introducing a simplified signal model. In this paper, we present a realistic block-based model and a practical design algorithm that handles real world signals. In doing so, we generalize known algorithms for multi-hypothesis MCP [2, 3].

Known AR models for multi-hypothesis MCP utilize Wiener coefficients to weight several hypotheses as well as different spatio-temporal positions. In contrast, we include a rate penalty to combine several spatio-temporally displaced hypotheses. For transmission, we assign relative spatio-temporal positions of hypotheses $(\mathbf{\Delta}_{x_\nu}, \mathbf{\Delta}_{y_\nu}, \mathbf{\Delta}_{t_\nu})$ to spatio-temporal displacement codewords. For the weighted superposition, all hypotheses are considered equally, independent of their spatio-temporal position. The predictor coefficients are not transmitted for each block.

This extension requires a rate-distortion framework. The average quality of the prediction has to be constrained by the average rate of the spatio-temporal displacement code. Section 2 explains rate-distortion optimized MCP. Section 3 introduces our model for block-based multi-hypothesis MCP. In section 4, we present the design algorithm, the optimal hypothesis selection algorithm, and the optimal predictor coefficients.

2 Rate-Distortion Optimized MCP

In block-based motion-compensated prediction, each block in the current frame is approximated by a spatially displaced block from the previous frame. We associate with each $s \times s$ block a vector in a s^2-dimensional space. Original blocks are represented by the random variable $\mathbf{S}$ with its samples $\mathbf{s}$ from the vector space.

The quality of the prediction is measured by the average distortion between original blocks $\mathbf{S}$ and predicted blocks $\hat{\mathbf{S}}$. We utilize squared Euclidean distance in the vector space to determine the distortion between two samples.

$$D = E\left\{\left\|\mathbf{S} - \hat{\mathbf{S}}\right\|_2^2\right\} \tag{1}$$

The blocks are coded with a displacement code $\mathbf{B}$. Each displacement codeword provides a unique rule how to compensate the current block-sample $\mathbf{s}$. The average rate of the displacement code is determined by its average length.

$$R = E\left\{|\mathbf{B}|\right\} \tag{2}$$

Optimal rate-distortion prediction minimizes average prediction distortion for a given average displacement rate. For our purposes, we restate the constrained problem. We weight the average rate by the Lagrange multiplier λ [5]. We call the resulting functional the *average rate-distortion measure* J. In order to achieve rate-distortion optimal prediction, we minimize the average rate-distortion measure for constant λ.

$$J(\lambda) = E\left\{\left\|\mathbf{S} - \hat{\mathbf{S}}\right\|_2^2\right\} + \lambda E\left\{|\mathbf{B}|\right\} \tag{3}$$

3 Block-Based Multi-Hypothesis MCP

Standard block-based motion-compensated prediction estimates one block from the previous frame in order to compensate one block in the current frame. We call this method short-term MCP. Long-term MCP as introduced in [4] extends the standard approach by estimating one block from several previous frames.

We introduce a new, block-based model for multi-hypothesis MCP. In contrast to short-term or long-term MCP, we use n blocks $\mathbf{c}_1, \ldots, \mathbf{c}_n$ from previous frames in order to predict one block in the current frame. All blocks which are available for prediction are called *hypotheses*. n hypotheses that predict the block $\hat{\mathbf{s}}$ are grouped to a *multi-hypothesis* or *n-hypothesis* $\mathbf{c}$. The predicted block $\hat{\mathbf{s}}$ is determined by linear combination of the individual components $\mathbf{c}_\nu$. The coefficients h_ν determine the weight of each component for the predicted block.

$$\hat{\mathbf{s}} = \sum_{\nu=1}^{n} \mathbf{c}_\nu h_\nu = \begin{pmatrix} \mathbf{c}_1 & \ldots & \mathbf{c}_n \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \mathbf{c}h \tag{4}$$

Figures 1 and 2 explain the difference between a 1-hypothesis and a 2-hypothesis for a two-dimensional block. The 1-hypothesis approximates the original block $\mathbf{s}$ directly. In contrast, the individual components of a 2-hypothesis do not necessarily approximate the original block; it is accomplished by combining the two components linearly.

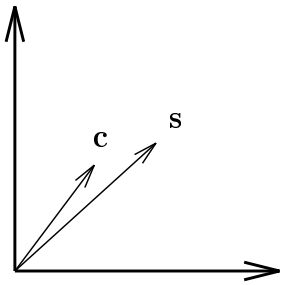

Figure 1: 1-hypothesis for a two-dimensional block.

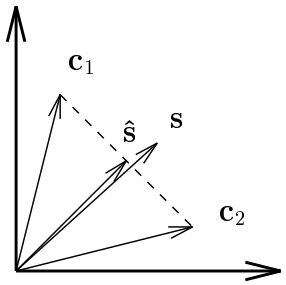

Figure 2: 2-hypothesis for a two-dimensional block.

The weighted superposition extends the predictive power of blocks available for prediction and causes also a dependence among the components of an n-hypothesis. Subsection 4.1 discusses this problem.

4 Block-Based Multi-Hypothesis MCP Design

We consider motion-compensated prediction as a vector quantization problem. For the design of multi-hypothesis MCP, we utilize known algorithms for vector quantizer design. The *Generalized Lloyd Algorithm* (GLA) in conjunction with *Entropy Constrained Vector Quantization* (ECVQ) solve the design problem iteratively.

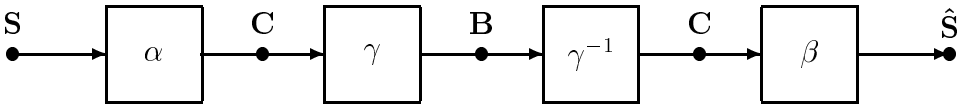

Figure 3: Model for the design of multi-hypothesis MCP.

The ECVQ algorithm [5] implies a quantizer model according to *Figure 3*, which we adopt for our design algorithm. The mapping α performs the estimation of the multi-hypotheses $\mathbf{C}$ from the original blocks $\mathbf{S}$. The mapping γ assigns each multi-hypothesis displacement to its corresponding entropy codeword. To be lossless, γ has to be invertible and uniquely decodable [5]. The weighted superposition β determines the predicted blocks $\hat{\mathbf{S}}$ from the multi-hypotheses.

For solving the design problem, we attempt to minimize the average rate-distortion measure (3) in order to get the optimal mappings α, β, and γ. The predicted blocks $\hat{\mathbf{S}} = \beta \circ \alpha(\mathbf{S})$ and the codewords $\mathbf{B} = \gamma \circ \alpha(\mathbf{S})$ can be expressed by our model mappings. The operator $\cdot \circ \cdot$ denotes composition according to *Figure 3*. This allows us to rewrite the average rate-distortion measure (3) in terms of the model and, in consequence, to determine the optimal predictor $\{\alpha, \beta, \gamma\}$.

$$J(\alpha, \beta, \gamma, \lambda, \mathbf{S}) = E\left\{\|\mathbf{S} - \beta \circ \alpha(\mathbf{S})\|_2^2 + \lambda|\gamma \circ \alpha(\mathbf{S})|\right\} \tag{5}$$

For given distribution of the original blocks $\mathbf{S}_c$ and constant Lagrange multiplier λ_c, the optimal predictor incorporates the optimal mappings α, β, and γ which satisfy

$$\min_{\alpha,\beta,\gamma} J(\alpha, \beta, \gamma, \lambda_c, \mathbf{S}_c). \tag{6}$$

Our iterative design algorithm for solving (6) includes three steps. The distribution of the original blocks $\mathbf{S}_c$ as well as the Lagrange multiplier λ_c are guessed for initialization.

The first step determines the optimal multi-hypothesis $\mathbf{c} = \alpha(\mathbf{s})$ for given mappings β_c and γ_c.

$$\begin{aligned} & \min_{\alpha} E\left\{\|\mathbf{S}_c - \beta_c \circ \alpha(\mathbf{S}_c)\|_2^2 + \lambda_c|\gamma_c \circ \alpha(\mathbf{S}_c)|\right\} \\ \Longrightarrow \quad & \alpha(\mathbf{s}) = \operatorname*{argmin}_{\mathbf{c}} \left\{\|\mathbf{s} - \mathbf{c}h_c\|_2^2 + \lambda_c|\gamma_c(\mathbf{c})|\right\} \end{aligned} \tag{7}$$

Equation (7) is the biased nearest neighbor condition familiar from vector quantization with a rate-constraint.

The second step provides the optimal mapping γ for given mappings α_c and γ_c. A constant mapping α_c assures a constant distribution of the multi-hypotheses $\mathbf{C}_c$.

$$\begin{aligned} & \min_{\gamma} E\left\{\|\mathbf{S}_c - \beta_c \circ \alpha_c(\mathbf{S}_c)\|_2^2 + \lambda_c|\gamma \circ \alpha_c(\mathbf{S}_c)|\right\} \\ \Longrightarrow \quad & \min_{\gamma} E\left\{|\gamma(\mathbf{C}_c)|\right\} \end{aligned} \tag{8}$$

Equation (8) postulates a minimum average codeword length for the optimal conditional code. For a finite number of multi-hypothesis displacements, the Huffman algorithm solves this optimization.

The third step determines the optimal multi-hypothesis superposition for given mappings α_c and γ_c.

$$\begin{aligned} & \min_{\beta} E\left\{\|\mathbf{S}_c - \beta \circ \alpha_c(\mathbf{S}_c)\|_2^2 + \lambda_c|\gamma_c \circ \alpha_c(\mathbf{S}_c)|\right\} \\ \Longrightarrow \quad & \min_{h} E\left\{\|\mathbf{S}_c - \mathbf{C}_c h\|_2^2\right\} \end{aligned} \tag{9}$$

Equation (9) is the Wiener problem for the optimal conditional predictor coefficients.

0: Given the distribution of the original blocks $\mathbf{S}$. Select the initial entropy code γ and predictor coefficients h. Choose a constant Lagrange mulitiplier λ.

1: Given the entropy code γ and predictor coefficients h, find the *optimal multi-hypothesis* $\mathbf{c}$ for each original block $\mathbf{s}$ according to

$$\min_{\mathbf{c}}\left\{\|\mathbf{s} - \mathbf{c}h\|_2^2 + \lambda|\gamma(\mathbf{c})|\right\}.$$

2: Given the new distribution of multi-hypotheses $\mathbf{C}$ from Step 1, design an *optimal entropy code* γ satisfying

$$\min_{\gamma} E\left\{|\gamma(\mathbf{C})|\right\}.$$

3: Given the multi-hypotheses $\mathbf{C}$ from Step 1, determine *optimal predictor coefficients* h by solving

$$\min_{h} E\left\{\|\mathbf{S} - \mathbf{C}h\|_2^2\right\}.$$

4: Continue with Steps 1 to 3 as long as improvement is possible.

Figure 4: Iterative design algorithm for multi-hypothesis MCP.

Figure 4 presents an iterative design algorithm based on the previous discussion. Step 2 incorporates an algorithm that minimizes the expected displacement codeword length and will not be discussed further. The conditional minimizations in Step 1 and Step 3 will be considered in more detail.

4.1 Optimal Hypothesis Selection Algorithm

According to our model, we have to find n hypotheses for each predicted block. The dependence among these hypotheses requires a joint solution for the estimation problem.

Each hypothesis is addressed by a spatio-temporal displacement $(\mathbf{\Delta}_{x_\nu}, \mathbf{\Delta}_{y_\nu}, \mathbf{\Delta}_{t_\nu})$. This address is relative to the position of the predicted block. Allowing a search space of size $[-a, a] \times [-a, a] \times [-m, -1]$, a full search algorithm implies a complexity of

$$P_f = \left[m(2a+1)^2\right]^n \tag{10}$$

search positions. For practical parameters ($a = 15$, $m = 10$, $n = 4$), the complexity of $P_f = 8.5 \cdot 10^{15}$ search positions is computationally too demanding.

An iterative algorithm, which is inspired by the *Iterated Conditional Modes (ICM)* of Besag [6], avoids searching the complete space by successively improving n optimal conditional solutions. Convergence to a local optimum is guaranteed, because the algorithm prohibits an increase of the error measure. A relative decrease of the rate-distortion measure of less than 0.5% indicates practical convergence. Our iterative version in *Figure 5* is called *Optimal Hypothesis Selection Algorithm (OHSA)* and provides a locally optimal solution for (7).

0: Assuming n-hypotheses, the rate-distortion measure

$$j(\mathbf{c}_1, \ldots, \mathbf{c}_\mu, \ldots, \mathbf{c}_n) = \left\| \mathbf{s} - \sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{n} \mathbf{c}_\nu h_\nu - \mathbf{c}_\mu h_\mu \right\|_2^2 + \lambda \left(\sum_{\substack{\nu=1 \\ \nu \neq \mu}}^{n} |\gamma(\mathbf{c}_\nu)| + |\gamma(\mathbf{c}_\mu)| \right)$$

is subject to minimization for each original block $\mathbf{s}$. Select the entropy code γ, predictor coefficients h, and the Lagrange multiplier λ. Initialize the algorithm with n hypotheses $(\mathbf{c}_1^{(0)}, \ldots, \mathbf{c}_n^{(0)})$ and set $i := 0$.

1: Select the μ-th out of n hypotheses; start from the first and end with the n-th hypothesis.

a: Focus on the μ-th hypothesis. All others are kept constant. Select a local neighborhood of hypothesis $\mathbf{c}_\mu^{(i)}$ as the conditional search space of hypothesis $\mathbf{c}_\mu^{(i+1)}$.

b: Minimize the rate-distortion measure by full search within the conditional search space of 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 measure decreases, continue with step 1 and set $i := i + 1$.

Figure 5: Optimal Hypothesis Selection Algorithm

Figure 6 demonstrates the performance for equally weighted hypotheses ($h_\nu = \frac{1}{n}$). Throughout the paper, we obtain our results by predicting from past frames of the original sequences. Prediction error is given as average PSNR in dB. Larger numbers indicate smaller prediction error variance. The results with half-pel accuracy are obtained by spatial bilinear interpolation.

We initialize the OHSA with n hypotheses by applying the rule of *Splitting One Hypothesis*. The computational demand of finding a 1-hypothesis is rather moderate. We repeat this optimal 1-hypothesis n times to generate the initial n-hypothesis.

For each n-hypothesis component in each iteration, OHSA performs a full search within a conditional search space in which an optimal conditional n-hypothesis com-

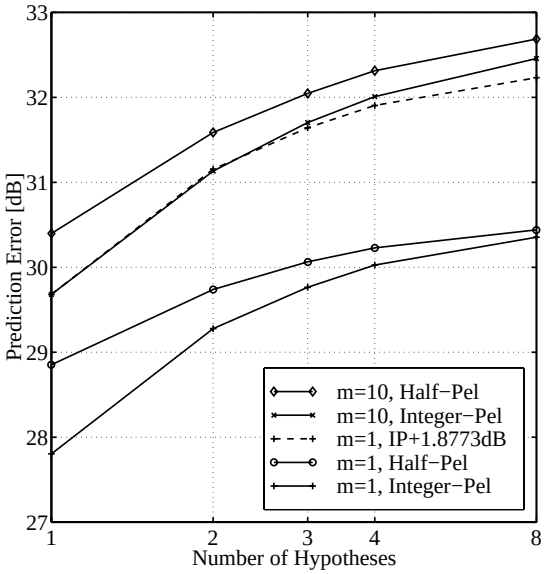

Figure 6: Prediction error and the number of hypotheses for the sequence *Foreman* (QCIF, 7.5 fps, 10s), 16×16 blocks, and $\lambda = 0$.

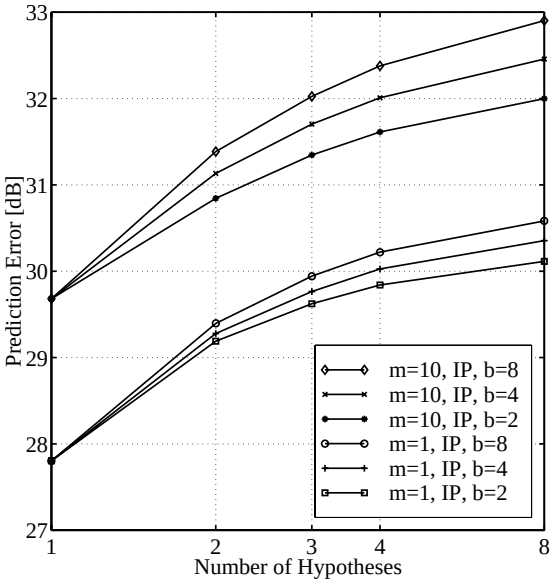

Figure 7: Prediction error and the number of hypotheses for the sequence *Foreman* (QCIF, 7.5 fps, 10s), integer-pel accuracy, 16×16 blocks, and $\lambda = 0$ dependent on conditional search space size.

ponent has to be found. The size of the conditional search space $[-b, b] \times [-b, b] \times [-b, b]$ affects the quality of the local optimum and the complexity of the algorithm, which is

$$P_i = m(2a+1)^2 + In(2b+1)^3 \tag{11}$$

search positions for I iterations. For practical parameters ($a = 15$, $m = 10$, $n = 4$, $b = 4$, $I = 3$), the complexity is reduced by factor $4.6 \cdot 10^{11}$ to $P_i = 1.8 \cdot 10^4$ search positions compared to (10). *Figure 7* shows the influence of the conditional search space size b.

OHSA does not determine the optimal number of hypotheses of a multi-hypothesis. The optimal number of hypotheses in the rate-distortion sense depends significantly on the rate constraint. For a given maximal number N, we determine the optimal number of hypotheses for each original block by running the OHSA for all numbers n from 1 to N and picking the one that minimizes the rate-distortion measure.

$$\min_{n:1 \le n \le N} \left\{ \left\| \mathbf{s} - \mathbf{c}^{(n)} h^{(n)} \right\|_2^2 + \lambda |\gamma(\mathbf{c}^{(n)})| \right\} \tag{12}$$

4.2 Optimal Predictor Coefficients

The third step in our iterative design algorithm solves the well-known Wiener problem of predictor design. Since our predictor preserves the expected value of the original

block, i.e. $E\{\mathbf{S}\} = E\left\{\hat{\mathbf{S}}\right\}$, we express the Wiener problem in covariance notation.

$$\min_{h} \left\{ C_{\mathbf{SS}} - 2h^T C_{\mathbf{CS}} + h^T C_{\mathbf{CC}} h \right\} \tag{13}$$

$C_{\mathbf{SS}}$ is the scalar variance of the original block, $C_{\mathbf{CC}}$ the $n \times n$ covariance matrix of the hypotheses, and $C_{\mathbf{CS}}$ the $n \times 1$ covariance vector between the hypotheses and the original block.

For video signals, we want to constrain additionally the sum of the prediction coefficients h_ν to one. With the vector $u^T = (1, 1, \cdots, 1)$ of dimension n, we write $u^T h = 1$. A Lagrangian approach to the constrained Wiener problem leads to the predictor coefficients

$$h = C_{\mathbf{CC}}^{-1} \left(C_{\mathbf{CS}} - \frac{u^T C_{\mathbf{CC}}^{-1} C_{\mathbf{CS}} - 1}{u^T C_{\mathbf{CC}}^{-1} u} u \right). \tag{14}$$

For the predictor design, we are using 18 training sequences each covering 10 seconds of video in QCIF resolution. The rate is 7.5 frames per second. The sequences are: *Akiyo, Bream, Car Phone, Children, Coastguard, Container Ship, Fun Fair, Hall Monitor, Mobile, News, Salesman, Sean, Silent, Stefan, Table Tennis, Total Destruction, Tunnel* und *Weather*.

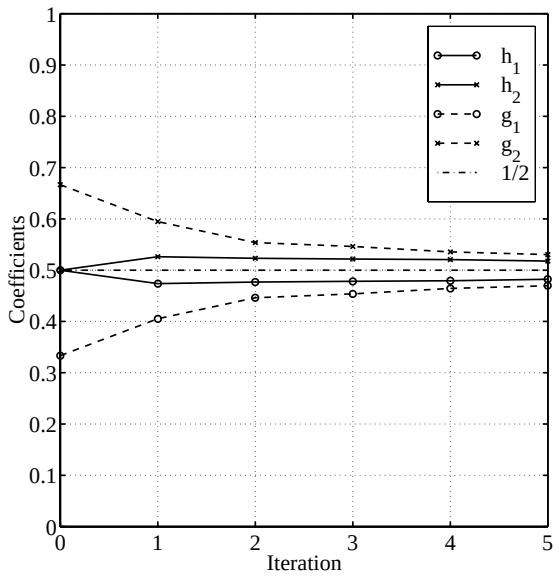

Figure 8: Convergence of the predictor coefficients (2 hypotheses) for the training sequences (QCIF, 7.5 fps), 16×16 blocks, $\lambda = 100$, and $m = 10$.

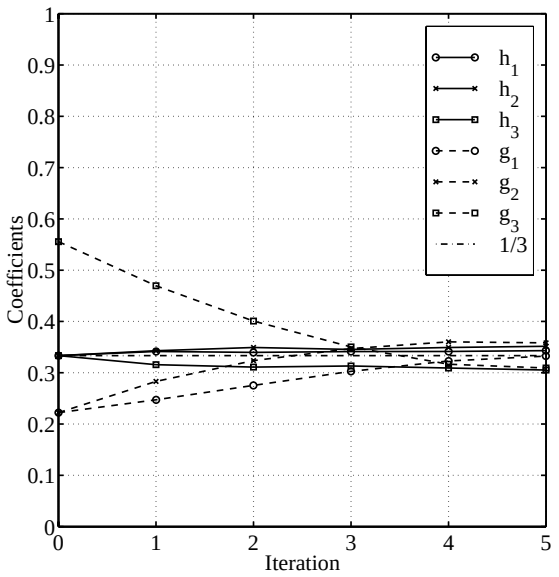

Figure 9: Convergence of the predictor coefficients (3 hypotheses) for the training sequences (QCIF, 7.5 fps), 16×16 blocks, $\lambda = 100$, and $m = 10$.

Figures 8 and 9 show the convergence of the predictor coefficients for an iterative predictor design. We use fixed length codebooks for initializing the design algorithm. In order to demonstrate convergence of predictor coefficients, we compare a uniform coefficient initialization $h_\nu = \frac{1}{n}$ to an arbitrary initialization g_ν. We observe that the converged predictor coefficients approximate $\frac{1}{n}$, regardless of their initial values.

4.3 Performance of Designed Predictors

We evaluate the rate-distortion performance for the designed predictors ($\lambda = 100$) by predicting the test sequence *Foreman* for various Lagrange multiplier $(25, 50, 100, \ldots, 1600)$. Note that the test sequence is not in the training set.

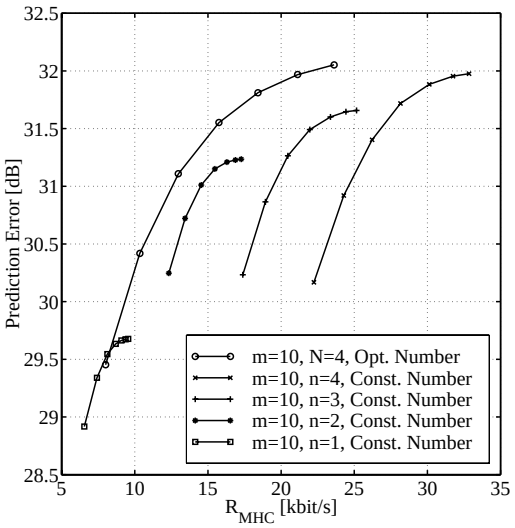

Figure 10: Prediction error and rate of the multi-hypothesis code R_{MHC} for the sequence *Foreman* (QCIF, 7.5 fps, 10s), and 16×16 blocks.

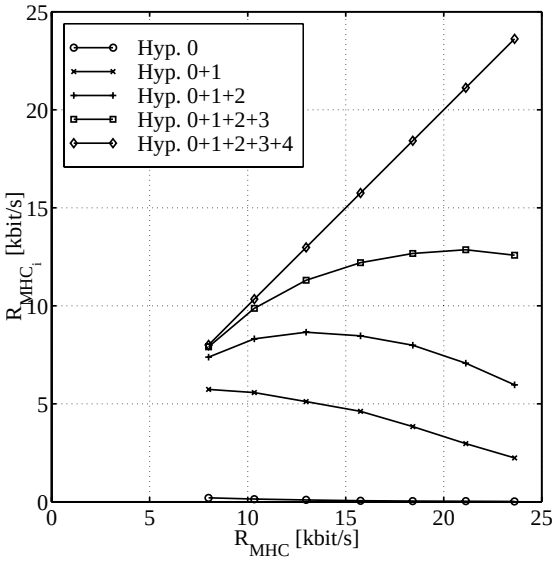

Figure 11: Cumulated partial rates R_{MHC_i} over the total rate R_{MHC} of the multi-hypothesis code for the sequence *Foreman* (QCIF, 7.5 fps, 10s), 16×16 blocks, $N = 4$, and $m = 10$.

Figure 10 compares 4 integer-pel predictors with a constant number of hypotheses to an adaptive integer-pel predictor allowing up to $N = 4$ hypotheses for each block. The case $n = 1$ corresponds to long-term MCP. Increasing the number of hypotheses from $n = 1$ to $n = 2$ provides gains of more than 1.5 dB in prediction error, the increase from $n = 1$ to $n = 4$ hypotheses, however, more than 2.2 dB. Please note the superiority of the adaptive predictor according to (12), which outperforms each predictor with constant number of hypotheses.

Figure 11 shows the subdivision of the total rate for the multi-hypothesis code R_{MHC} generated by the adaptive predictor. Hypothesis "0" marks an uncoded block and is a special case of a 1-hypothesis. Partial rates of the multi-hypothesis code correspond to differences between two successive curves in *Figure 11*. We observe that more blocks are decomposed into $N = 4$ hypotheses and less blocks into $n < N$ hypotheses by increasing the rate.

Multi-hypothesis motion compensated prediction is a very promising technique that can yield significant bit-rate savings for future video coding algorithms. Increasing the accuracy of MCP from integer-pel to half-pel provides gains from 0.7 dB to 1 dB in prediction error (*Figure 6*). But increasing the number of hypotheses from $n = 1$ to $n = 4$ provides gains of more than 2.2 dB in prediction error.

5 Conclusions

In this paper, we have presented a locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction. We explicitly distinguish between the search space and the superposition of hypotheses. The components of a multi-hypothesis are selected from the same search space and their spatio-temporal positions are transmitted by means of spatio-temporal displacement codewords. Constant predictor coefficients are used to combine linearly components of a multi-hypothesis. Further, we have provided an estimation criterion for optimal n-hypotheses, a rule for optimal displacement codes, and a condition for optimal predictor coefficients. Statistically dependent components of a n-hypothesis are determined by the optimal hypothesis selection algorithm, which improves successively n optimal conditional hypotheses. For best performance, we additionally determine the optimal number of hypotheses for each original block.

Several important observations can be made. Increasing the number of hypotheses from 1 to 2 provides prediction gains of more than 1.5 dB, the increase from 1 to 4 hypotheses more than 2.2 dB. OHSA reduces the complexity of the underlying joint optimization problem to a feasible size. Determining the optimal number of hypotheses for each block, additional improvements are achieved. For increasing rate constraint, the average number of hypotheses decreases for each original block. Finally, we observe practically no dominant n-hypothesis component for our training sequences. The optimum predictor coefficients are approximately $\frac{1}{n}$ for n linear combined hypotheses.

References

[1] B. Girod, "Efficiency Analysis of Multi-Hypothesis Motion-Compensated Prediction for Video Coding", Submitted to IEEE Trans. on Image Processing, 1997.

[2] M.T. Orchard and G.J. Sullivan, "Overlapped Block Motion Compensation: An Estimation-Theoretic Approach", *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 693–699, Sept. 1994.

[3] S.-W. Wu and A. Gersho, "Joint Estimation of Forward and Backward Motion Vectors for Interpolative Prediction of Video", *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 684–687, Sept. 1994.

[4] T. Wiegand, X. Zhang, and B. Girod, "Motion-Compensating Long-Term Memory Prediction", in *Proc. of the IEEE Int. Conf. on Image Processing*, 1997.

[5] P.A. Chou, T. Lookabaugh, and R.M. Gray, "Entropy-Constrained Vector Quantization", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, pp. 31–42, Jan. 1989.

[6] J. Besag, "On the Statistical Analysis of Dirty Pictures", *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.