Still Image Compression, Proc. SPIE 24i8, pp. 199–207, Feb. 1995

Image Subband-Coding Using an Information-Theoretic Subband Splitting Criterion

Ulug Bayazit and William A. Pearlman
Electrical, Computer and Systems Engineering Department
Rensselaer Polytechnic Institute, Troy, New York

May 23, 1996

ABSTRACT

It has been proved recently that for Gaussian sources with memory an ideal subband split will produce a coding gain for scalar or vector quantization of the subbands. Following the methodology of the proofs, we outline a method for successively splitting the subbands of a source, one at a time to obtain the largest coding gain. The subband with the largest *theoretical rate reduction* (TRR) is determined and split at each step of the decomposition process. The TRR is the difference between the rate in optimal encoding of N-tuples from a Gaussian source (or subband) and the rate for the same encoding of its subband decomposition.

The TRR is a monotone increasing function of a so-called spectral *flatness ratio*, which involves the products of the eigenvalues of the source (subband) and subband decomposition covariance matrices of order N. These eigenvalues are estimated by the variances of the Discrete Cosine Transform, which approximates those of the optimal Karhunen Loeve Transform.

After the subband decomposition hierarchy or tree is determined through the criterion of maximal TRR, each subband is encoded with a variable rate entropy constrained vector quantizer. Optimal rate allocation to subbands is done with the BFOS algorithm which does not require any source modelling.

We demonstrate the benefit of using the criterion by comparing coding results on a two-level low-pass pyramidal decomposition with coding results on a two-level decomposition obtained using the criterion. For 60 MCFD (Motion Compensated Frame Difference) frames of the Salesman sequence an average rate-distortion advantage of 0.73dB and 0.02bpp and for 30 FD (Frame Difference) frames of Caltrain image sequence an average rate-distortion advantage of 0.41dB and 0.013bpp are obtained with the optimal decomposition over low-pass pyramidal decomposition.

Keywords : subband decomposition, splitting, vector quantization, image sequence coding

1 INTRODUCTION

Recently it has been proved by Rao and Pearlman 4 that greater entropy reduction can be achieved by finite order prediction from each of the subbands of a Gaussian source than by the same order prediction of the source itself. It is then clear that optimum block or block transform coding in the rate-distortion sense will yield better performance on the subbands than on the fullband provided that the coding distortion is kept low. Then the question is how to optimally perform a frequency decomposition that will give us the greatest gain in the operational rate-distortion performance. Although the general problem of decomposition has several subproblems of interest which should be separately addressed, such as which subband to split, number of new subbands resulting from each decomposition, bandwidth of each subband etc., we shall limit our attention to the simpler problem of dyadic tree decomposition which sufficiently solves the first two of these subproblems.

As in many of the tree-structured search algorithms present in the literature of source coding, the subband decomposition criterion of this paper is a single step lookahead type of search algorithm which evaluates the immediate effects of the split of each subband before it decides on which subband should be split. A similar single step lookahead type of algorithm was suggested by Coifman and Wickerhauser in 1 for the pruning of a library of orthonormal wavelet bases based on the criterion function of entropy of the subband coefficient magnitudes. From our perspective it is not quite clear how this criterion function relates to rate-distortion theoretic considerations. In 4, Rao and Pearlman have also looked at a no step lookahead type of algorithm using the spectral flatness of each subband as the criterion function. Spectral flatness is a measure of the uncorrelatedness of a source. However it is not the only determinant of the rate-distortion performance of a Gaussian source. Variance of the Gaussian source plays an equally important role. It is also worth noting from the rate-distortion characteristics of a Gaussian source that the flatter the source spectrum is the larger the number of bits it requires to achieve a given signal-to-noise-ratio. A less correlated source may not be significantly further decorrelated by an orthogonal tranformation such as subband decomposition and hence will yield a smaller gain percentagewise than a more correlated source. On the other hand, a less correlated source may require so many more bits than a correlated source that the gain in rate due to decomposition of the less correlated source can be significant and even exceed that due to decomposition of the more correlated source. This intuitive explanation is the motivation for considering high-pass as well as low-pass bands for splitting at each step of the decomposition process.

A rough estimation of N'th order spectral flatness may be obtained by computing the variance of the optimum Karhunen-Loeve Transform (KLT) coefficients. Since KLT transform is signal dependent, coefficients of the Discrete Cosine Transform (DCT) can be substituted for this purpose. There is no additional computational complexity required by the actual coding process if DCT is also used to code the subband transform coefficients. A block transform approach for coding the subbands is valid because subbands pixels have mutual correlation. In this sense a spatial block coding technique such as vector quantization may be found redundant in a practical system. However since our goal here is to show the rate-distortion superiority of the new decomposition criterion rather than design a practical system, Entropy-Constrained Pairwise Nearest Neighbor (ECPNN) an efficient multi-rate vector quantization algorithm developed by de Garrido et. al. in 6 has been used for the purpose of block coding system design. Optimal rate allocation to subbands is done with the BFOS algorithm which uses the operational rate-distortion characteristics of the subbands to arrive to the distribution of rate yielding the least distortion.

A possible application of this decomposition technique is to adaptively decompose the FD (Frame Difference) or MCFD (Motion compensated frame difference) frames of an image sequence. A map of the tree structure used for subband decomposition may be sent as side information for each frame.

In Section 2 the Gaussian model which our work assumes is introduced together with notation to be used throughout this paper. Section 3 gives an explanation of why rate-distortion theoretic improvement is to be expected through subband coding based on our Gaussian model. Section 4 presents the derivation of the splitting criterion used for the decomposition. Section 5 discusses one of the potential problems involved with the splitting criterion due to the assumption of ideal brickwall filters made in its derivation. Section 6 briefly covers the coding system employed to test the performance of the splitting criterion. Section 7 presents results on image sequences obtained by using the criterion and compares them with conventional low-pass pyramid coding results for a two level (seven subband) hierarchy. Section 8 summarizes the important points of this paper along with some of the shortcomings of our approach.

2 Preliminaries

For Gaussian sources, $X = (X_0, X_1, \ldots)$, the N-th order differential entropy per source sample is given by

$$h_N(X) = \frac{1}{2} log_2(2\pi e \gamma_{N,X}^2 \sigma_X^2) \tag{1}$$

where $\gamma_{N,X}^2$ and σ_X^2 are spectral flatness and variance of the source X.

$$\gamma_{N,X}^2 = \frac{\prod_{i=1}^{N}(\lambda_i)^{\frac{1}{N}}}{\sigma_X^2} \tag{2}$$

$$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}\lambda_i \tag{3}$$

λ_i is the i-th eigenvalue of the N dimensional covariance matrix of the stationary Gaussian source. In other words λ_i's are the variances of the discrete Karhunen Loeve Transform (KLT) coefficients.

Note that from the non-negativity of the geometric-to-arithmetic mean inequality we obtain

$$0 \le \gamma_{N,X}^2 \le 1 \tag{4}$$

with equality to one if and only if the source spectrum is flat, i.e., source is white.

The entropy rate is related to Nth order differential entropy as

$$h_\infty(X) = \lim_{N\to\infty} h_N(X) = \frac{1}{2} log_2(Q_X) \tag{5}$$

where $Q_X = exp\{\frac{1}{2\pi}\int_{-\pi}^{\pi} log_2 S_X(\omega) d\omega\}$ is the source entropy power. Following these definitions the average mutual information between N members of the sequence and the rest of the sequence may be defined as

$$\begin{aligned} {}^N\Delta_\infty^X &= \lim_{K\to\infty} I(X_0, X_1, \ldots, X_{N-1}; X_N, \ldots, X_K) \\ &= h_N(X) - h_\infty(X) \end{aligned} \tag{6}$$

When the source is split into subbands and subsampled, the composite subband average mutual information may also be defined as

$${}^N\Delta_\infty^{SB} = \sum_{i=1}^{M} V_{SBi}\, {}^N\Delta_\infty^{SBi} \tag{7}$$

where ${}^{N}\Delta_{\infty}^{SBi}$ denotes the subband mutual information as in Eqn. (6) and V_{SBi} is the inverse of the subsampling factor.

The N'th order rate distortion function for a Gaussian source X for low distortions is given as

$$R_{N,X}(D_X) = \frac{1}{2} log_2(\frac{\gamma_{N,X}^2 \sigma_X^2}{D_X}) \tag{8}$$

where $R_{N,X}$ denotes rate in bits per sample of source X and D_X denotes mean squared distortion per sample of source X.

Since the KLT transform has no fast algorithm in general and is signal dependent, suboptimal DCT is usually substituted. We may replace λ_i in Eqn.(2) by the variances of DCT coefficients. This is well justified because the correlation between different coefficients of the DCT is quite small for most image sources and these variances converge to the λ_i's in the limit of large N. Moreover, the DCT coefficients converge in distribution to a Gaussian process in this limit. We further replaced the variances by sample variances since for each subband we had a large number of DCT blocks at our disposal. In cases where the available number of DCT blocks is small a method such as maximum entropy spectral estimation must be used.

The derivation in the next two sections makes the assumption of ideal brickwall subband filters. These filters are only of theoretical interest and are not realizable. The implication of this assumption is discussed in Section 5.

3 Rate Distortion Theoretic Improvement Due to Subband Coding

In this section we restate some of the recently proved important theorems justifying the importance of block coding of subbands.

3.1 Theorem4

For Gaussian AR sources, the composite N'th order entropy of the subbands, $h_N(SB)$, is closer to the entropy rate of the source than the N'th order entropy of the fullband X. i.e.,

$$\Delta H =^N \Delta_{\infty}^{X} -^N \Delta_{\infty}^{SB} \geq 0. \tag{9}$$

Because the entropy rate, $h_{\infty}(X)$, is unchanged by an ideal subband decomposition, subband decomposition results in a decrease of the N'th order entropy of the source. It remains to show that combined N'th order rate distortion function of the subbands constitutes a lower bound to the N'th order rate distortion function of the fullband. This follows from the theorem stated above.

3.2 Theorem 2, 5

For Gaussian AR sources, the combined N'th order operational rate distortion function of the subbands constitutes a lower bound to the N'th order rate distortion function of the fullband if distortion is low.

$$R_{N,SB}(D_X) \leq R_{N,X}(D_X) \quad if \;\; \lambda_i \geq D_X \;\; for \;\; i = 1, \cdots, N \tag{10}$$

Proof: From the above theorem we know that $h_N(X) \geq h_N(SB)$. From 2, Eqn. (42)

$$\begin{aligned} R_{N,X}(D_X) - R_X(D_X) &\leq \frac{1}{2} log_2 \frac{\gamma^2_{N,X}\sigma^2_X}{Q_X} \\ &= h_N(X) - h_\infty(X) =^N \Delta^X_\infty \end{aligned} \tag{11}$$

with equality if $\lambda_i \geq D_X \;\; \forall i$. For $\lambda_i \geq D_X \;\; \forall i$ we obtain

$$\begin{aligned} R_{N,X}(D_X) - R_{N,SB}(D_X) &= {}^N\Delta^X_\infty -^N \Delta^{SB}_\infty \\ &= \Delta H \geq 0 \end{aligned} \tag{12}$$

and the claim of the theorem follows from the previous theorem.

The combined N'th order rate-distortion function of the subband process is

$$\begin{aligned} R_{N,SB}(D_X) &= \sum_{i=1}^{M} V_{SBi} R_{N,SBi}(D_{SBi}) \\ &= \frac{1}{2} log_2 \frac{\prod_{i=1}^{M}(\gamma^2_{N,SBi}\sigma^2_{SBi})^{V_{SBi}}}{\prod_{i=1}^{M}(D_{SBi})^{V_{SBi}}} \\ &= \frac{1}{2} log_2 \frac{\prod_{i=1}^{M}(\gamma^2_{N,SBi}\sigma^2_{SBi})^{V_{SBi}}}{\prod_{i=1}^{M}(V_{SBi})^{V_{SBi}} D_X} \end{aligned} \tag{13}$$

The substitution in the second equality $D_{SBi} = V_{SBi} D_X$ will be justified in the next section.

4 Criterion for Maximum Rate Distortion Theoretic Gain

Before making the substitution $D_{SBi} = V_{SBi} D_X$ we can write the rate reduction due to subband coding as

$$\begin{aligned} \Delta R_{N,X}(D_X, D_{SBi}) &= R_{N,X}(D_X) - \sum_{i=1}^{M} V_{SBi} R_{N,SBi}(D_{SBi}) \\ &= \frac{1}{2}\log_2\left(\frac{\gamma^2_{N,X}\sigma^2_X}{\prod_{i=1}^{M}(\gamma^2_{N,SBi}\sigma^2_{SBi})^{V_{SBi}}} \frac{\prod_{i=1}^{M}(D_{SBi})^{V_{SBi}}}{D_X}\right) \end{aligned} \tag{14}$$

When equating the fullband coding distortion and the distortion from reconstructed subbands

$$D_X = \sum_{i=1}^{M} D_{SBi} \tag{15}$$

Eqn. (14) can be rewritten as

$$\Delta R_{N,X}(D_{SBi}) = \frac{1}{2}\log_2\left(\frac{\gamma^2_{N,X}\sigma^2_X \prod_{i=1}^{M}(V_{SBi})^{V_{SBi}}}{\prod_{i=1}^{M}(\gamma^2_{N,SBi}\sigma^2_{SBi})^{V_{SBi}}} \frac{\prod_{i=1}^{M}(\frac{1}{V_{SBi}} D_{SBi})^{V_{SBi}}}{\sum_{i=1}^{M} D_{SBi}}\right) \tag{16}$$

By using geometric-arithmetic mean inequality for the ratio of the terms involving distortions we see that

$$\Delta R_{N,X}(D_{SBi}) \leq \frac{1}{2}\log_2\left(\frac{\gamma^2_{N,X}\sigma^2_X \prod_{i=1}^{M}(V_{SBi})^{V_{SBi}}}{\prod_{i=1}^{M}(\gamma^2_{N,SBi}\sigma^2_{SBi})^{V_{SBi}}}\right) \tag{17}$$

with equality if and only if $D_{SBi} = V_{SBi} D_X$. Note that to maximize the gain on the left handside of Eqn. (16) we should independently maximize the two ratios involving entropy rates (product of spectral flatness and variance) and distortions. The first of these maximization processes depends on the subband decomposition of the source and the second one depends on rate allocation to subbands. Assumption of optimal rate allocation to subbands implies $D_{SBi} = V_{SBi} D_X$.

Let's define the *flatness ratio*, FR, as

$$FR = \frac{\gamma_{N,X}^2 \sigma_X^2 \prod_{i=1}^{M} (V_{SBi})^{V_{SBi}}}{\prod_{i=1}^{M} (\gamma_{N,SBi}^2 \sigma_{SBi}^2)^{V_{SBi}}} \tag{18}$$

FR is related to the Theoretical Rate Reduction (TRR) by

$$TRR = \frac{1}{2} \log_2(FR) \tag{19}$$

TRR is the maximum theoretically achievable rate reduction for coding N-tuples of the subband decomposition over coding N-tuples of the nondecomposed source.

At each step of a successive subband splitting process one can pick the subband with the maximum FR or TRR to be split under the assumption of optimal rate allocation to follow. This is a one-step lookahead criterion since it requires the variance and spectral-flatness of the subbands of each one of the candidate subbands to be computed before a decision is made as to which candidate subband should actually be split.

In this paper we consider splits which divide the two-dimensional first-level frequency subband in question into four equal bandwidth subbands. Hence $V_{SBi} = \frac{1}{4}$ for our purposes.

5 Performance Loss due to QMF Filters and Prediction Threshold

Up to this point we were interested in devising a criterion which would inform us which of the candidate subbands should be split if a split is to be performed. The question of whether or not splitting should be performed indefinitely or should be stopped after a certain stage in the overall decomposition process deserves some attention. As noted in 3 and later in 5 the qmf nature of actual subband filters results in rate-distortion theoretic loss in the limit of large block sizes (see 3) and also for block sizes greater than the *prediction threshold* (see 5 for a definition). We conjecture here that the *prediction threshold* decreases as we successively decompose the spectrum. As a result after a certain number of successive splits we can expect a performance loss even with a small dimensional block coding scheme since the threshold will be smaller than the block dimension. This phenomenon is demonstrated with our simulations in Section 7.

6 Coding System for Testing the Splitting Criterion

After the splitting hierarchy is determined through the TRR criterion and the desired decomposition structure is obtained, each resulting subband is block coded with a multi-rate vector quantizer. ECPNN, 6, the algorithm used to create the vector quantizer, designs a codebook of a desired rate from an initial large size codebook by repeatedly merging two codevectors at a time. Each merge results in a smaller size codebook with a lower rate and higher distortion. The two codevectors are chosen such that the ratio of the increase in distortion to the decrease in rate is minimized. Rate here is defined as the entropy of the codevector indices. The initial large size codebook is designed by the LBG algorithm. The algorithm designs vector quantizers with fairly decent operational rate-distortion characteristics which closely approximate those of ECVQ.

After the desired overall coding rate is specified, each subband vector quantizer is allocated rate using the BFOS algorithm. Starting from the initial codebook each ECPNN merge results in a slightly different codebook as explained above. The pair of rate distortion values of each codebook is registered as an element of a sequence for the particular subband. A large tree is constructed with its root node having one child per subband. The subtree rooted at each child is then the degenerate unary tree corresponding to the linearly linked sequence of codebooks for each subband. The root node of each unary tree is the zero rate codebook and the bottommost node is the maximum size codebook for each subband. The large tree is pruned to a desired rate using the criterion of achieving the least increase in overall distortion with respect to a decrease in overall rate. It can be shown that at each pruning step there exists a subtree with its root coinciding with the root of the original tree which satisfies this criterion. At the end of each pruning step the overall tree loses a part of one of its unary trees and a smaller tree improved in the rate-distortion sense is generated. The unary tree may be pruned by more than one node at a time. Successive prunings using this criterion leads to the tracing of the convex hull of the operational rate-distortion function.

In the encoding stage each subband source vector is encoded with the codevector of the subband vector quantizer which minimizes the mean squared distortion plus a Lagrange multiplier times the source codeword length (hence the term entropy-constrained). The Lagrange multiplier is given by the slope of the distortion-rate characteristics of the subband vector quantizer at the operating point.

7 Results

The criterion presented above has been used to find better frequency decomposition structures than the low-pass pyramid one commonly encountered in literature for FD (Frame Difference) and MCFD (Motion Compensated Frame Difference) frames of image sequences. Before the actual coding of each frame, the optimal decomposition structure can be determined by repeatedly splitting using the above criterion. Each of the resulting subbands is allocated rate by the BFOS algorithm and block coded by the multi-rate quantizer as explained in the previous section.

To keep the simulations simple the following design considerations have been used. After an initial mandatory split of the fullband frame, only one of the resulting four first level subbands (LL, LH, HL, HH) is split using the criterion to form a two-level hierarchy. Although the DCT variance approximation to the KLT eigenvalues may not be as close for small vector dimensions, dimensions as small as 2x2 have been used for all bands. Note that a vector dimension of 1x1 corresponds to scalar quantization of each subband and spectral flatness reduces to 1. The criterion in this case is proportional to a ratio involving variances only.

The simulation called the optimum decomposition is the one in which the first level subband yielding the highest theoretical rate reduction with the splitting criterion is split to obtain the two-level hierarchy. For comparision with a two level low-pass pyramid the LL band is also split and the resulting simulation is called the suboptimum decomposition. Further parameters of interest are block matching blocksize of 16x16, search area size of 15x15, full pixel accuracy with forward motion estimation. Adelson Simoncelli 9-tap qmf filters were used. In all the coding simulations the coded sequence is the sequence from which the training set is obtained.

The TRR values for the splitting of each one of the four first level subbands of the first thirty FD frames of the Caltrain sequence (360x288x30) are plotted in Figure 1. The sequence has a frame rate of 30Hz. From Figure 1 the optimum decomposition is the splitting of the HL band for all the FD frames. Coding results for these FD frames are given in Figures 3 and 4. An average coding gain of 0.41dB and 0.02bpp. is obtained by splitting the HL band over splitting the LL band. For this simulation the training set is six of the FD frames of the original sequence sampled at 3.75 Hz.

The TRR values are also given for the first sixty MCFD frames of Salesman sequence (360x288x60) in Figure 2. In this case the TRR criterion determines that splitting of the HH band band will yield the greatest rate reduction. Coding results for these MCFD frames are presented in Figures 5 and 6. An average gain of 0.73dB and 0.02bpp. is obtained by splitting the HH band over splitting the LL band. The training set for this simulation set is six of the MCFD frames sampled at 3.0 Hz from the original sequence. In both simulation sets the coding results for the optimum decomposition are significantly better than the coding results for the suboptimum decomposition. These results corroborate our theoretical rate reduction predictions.

Note that although the TRR values obtained are always positive, splitting does not necessarily improve performance in practice. In Figures 3 and 4 the suboptimum decomposition of the LL band yields inferior results to that of a one-level hierarchy and the optimum decomposition does not do significantly better than the one-level hierarchy. It is likely that these subbands are so spectrally flat or uncorrealated that the QMF filters used do not allow operation above the low prediction threshold. However we have not observed this phenomenon for most of the simulations that we have performed. Nevertheless it brings about the necessity for further research for a more accurate QMF based splitting criterion which can take into account non-ideal filters and can inform the user of performance gain or loss to be expected with further splitting. Nevertheless, the ideal filtering based criterion in its present form is able to rank the candidate subbands in rate reduction expected in a hierarchical splitting procedure and hence reliably point to which subband split will yield the greatest rate reduction.

The time required for computing the criterion for the candidate subbands of each frame is relatively small compared to the overall processing time required to encode each frame. In our simulations for the image sequences, computation of the criterion for the candidate subbands of each frame took only about five to ten CPU seconds while the overall processing time for each frame was of the order of several minutes.

8 Conclusion

A subband splitting criterion has been presented based on the assumptions that the source is Gaussian, the distortion is low and the subband filters are ideal for subsequent block coding. Preliminary results obtained by using this criterion to generate a two level pyramidal hierarchy for the FD and MCFD frames of image sequences are encouraging. One reason for questioning the reliability of this criterion comes from the fact that the actual filters used are QMF, not ideal filters. The criterion cannot predict a performance loss due to the aliasing of coding distortions with QMF filters if large block dimensions are used. Another problem arises from the DCT approximation to the KLT transform used to estimate the eigenvalues since the block size is small. Yet another problem is the low-rates at which the coding system is operating. This inevitably introduces high distortion and might contradict the initial assumption made to obtain the simplified Gaussian model. Although none of these assumptions is precisely true, the ranking of rate reduction among subbands predicted by the criterion is conjectured to hold true for real world sources and filter banks. Finally the application of the criterion does not significantly increase complexity of the coding system.

REFERENCES

1. R. F. Coifman, M. V. Wickerhauser, "Entropy-Based Algorithms for Best Basis Selection," *IEEE Transactions on Information Theory,* vol. 38, no. 2, March 1992.

2. S. Nanda, W. A. Pearlman, "Tree Coding of Image Subbands,' ' *IEEE Transactions on Image Processing,* vol. 1, no. 2, April 1992.

3. T. R. Fischer, "On the Rate-Distortion Efficiency of Subband Coding," *IEEE Transactions on Information Theory,* vol. 38, no. 2, March 1992.

4. S. Rao, W. A. Pearlman, "Analysis of Linear Prediction, Coding and Spectral Estimation from Subbands," *IEEE Transactions on Information Theory,* accepted for publication.

5. S. Rao, "Analysis of Linear Prediction,Coding and Spectral Estimation from Subbands," Ph.D. Dissertation, Rensselaer Polytechnic Institute, Troy, NY 1994.

6. D. P. de Garrido, W. A. Pearlman and W. A. Finamore,' 'A clustering algorithm for entropy-constrained vector quantizer design with applications in coding image pyramids", *IEEE Transactions on Circuits and Systems for Video Technology,* submitted.

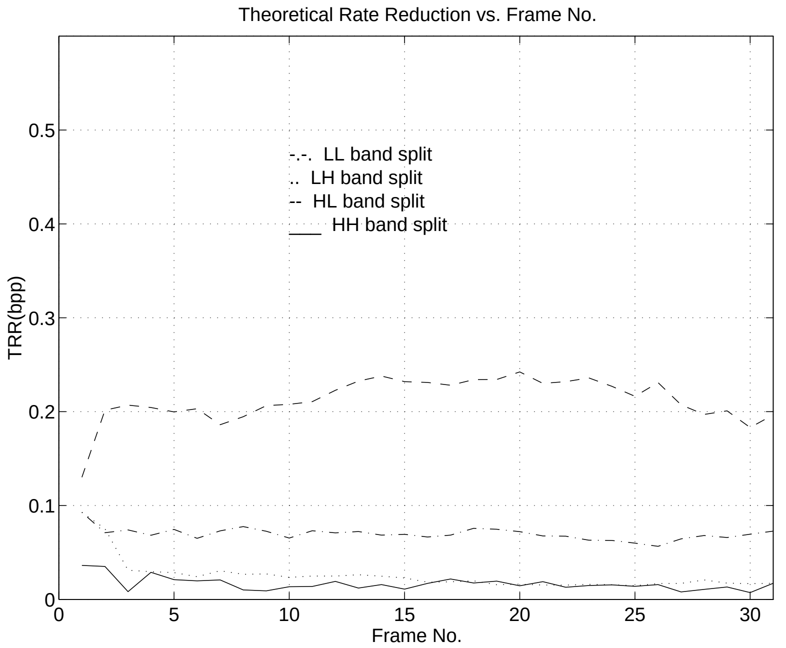

Figure 1: Application of the splitting criterion to the FD frames of Caltrain sequence

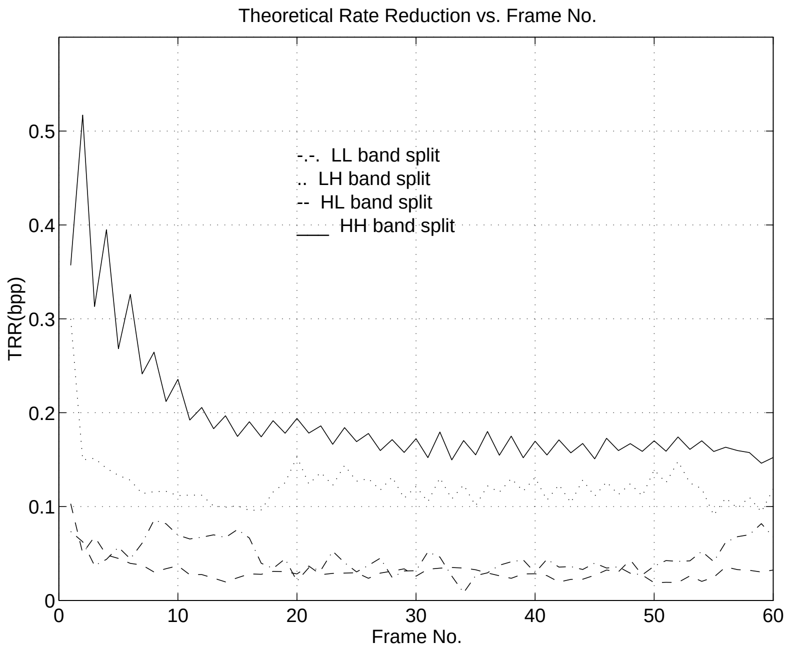

Figure 2: Application of the splitting criterion to the MCFD frames of Salesman sequence

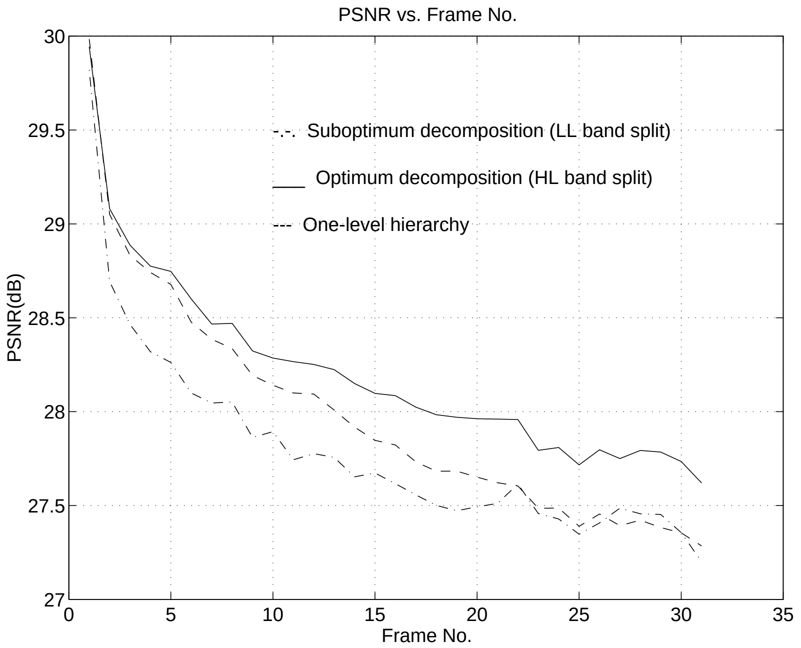

Figure 3: Simulation results with Caltrain sequence

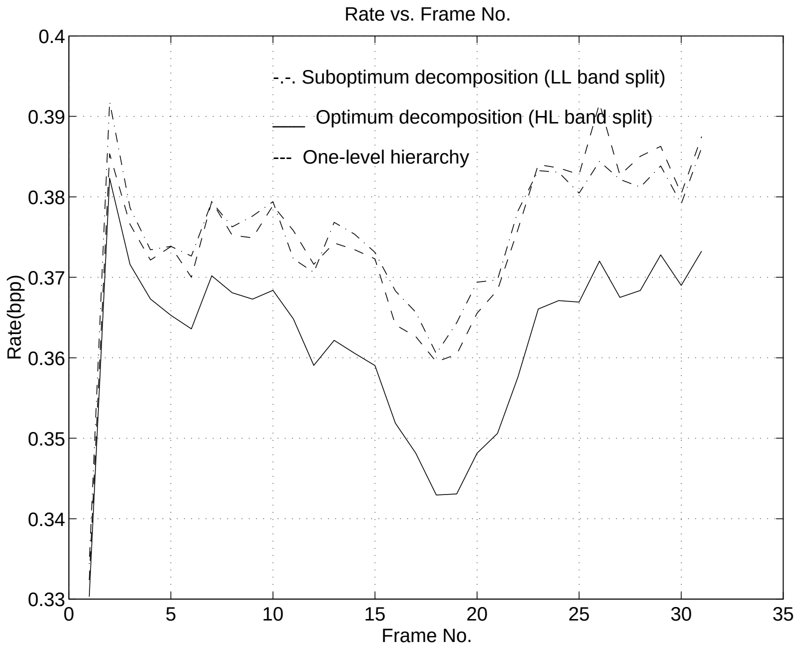

Figure 4: Simulation results with Caltrain sequence(cont.)

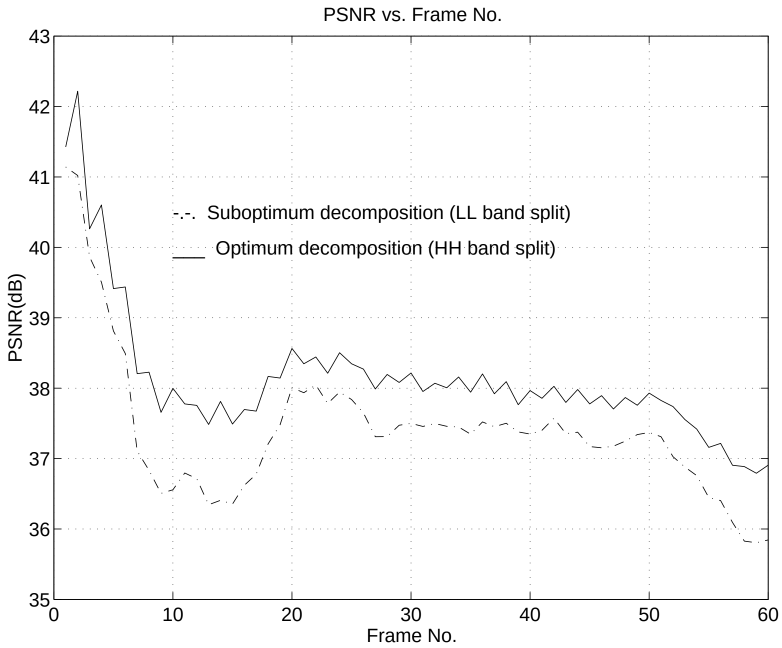

Figure 5: Simulation results with Salesman sequence

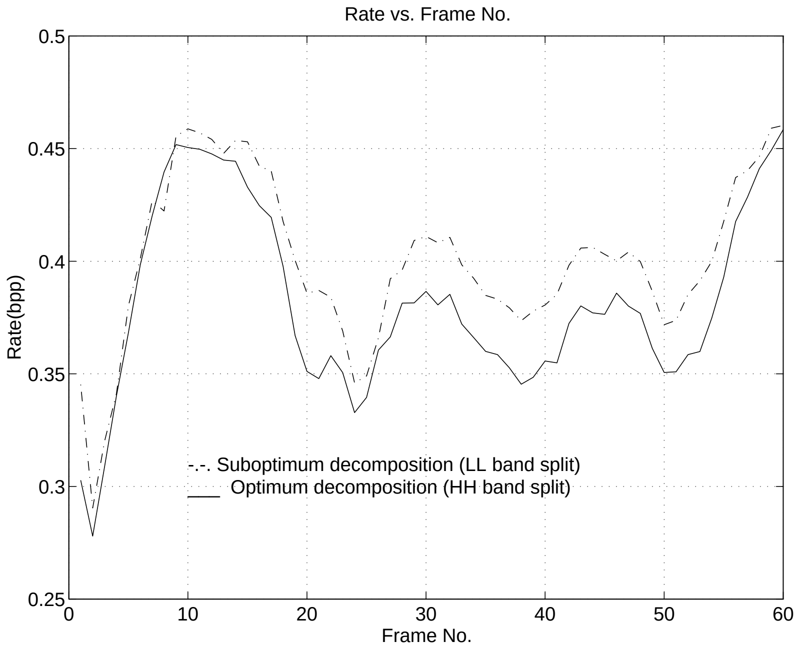

Figure 6: Simulation results with Salesman sequence(cont.)