VIDEO CODING WITH MOTION COMPENSATION FOR GROUPS OF PICTURES

Markus Flierl and Bernd Girod

Information Systems Laboratory
Stanford University, Stanford, CA 94305
{mflierl,bgirod}@stanford.edu

ABSTRACT

This paper discusses the efficiency of a compression scheme for video sequences that jointly encodes groups of pictures. Our approach, motion-compensated transform coding, applies a KLT to decorrelate a set of motion-compensated pictures for efficient encoding. The theoretical investigation utilizes a signal model for inaccurate motion compensation and provides a performance comparison to motion-compensated prediction. We discuss the influence of motion accuracy, residual noise, and the correlation of displacement errors dependent on the number of coded pictures.

1. INTRODUCTION

Today's video coding schemes utilize DPCM with motion-compensated prediction (MCP) for efficient compression. Such compression schemes require sequential processing of video signals which makes it difficult to achieve efficient embedded representations of video sequences. This paper investigates the efficiency of motion-compensated transform coding, a compression scheme which jointly encodes groups of pictures (GOP). For that, we utilize a powerful model for inaccurate motion compensation with additive residual noise. This model has proven to be useful to characterize motion-compensated prediction [1] and multihypothesis motion-compensated prediction [2]. Our approach for jointly encoding groups of pictures applies the KLT to decorrelate a set of motion-compensated pictures. To evaluate the performance of motion-compensated transform coding, we assume compression at high bit-rates and determine the rate difference to optimum intraframe coding of individual motion-compensated pictures. In the following, Section 2 introduces the model for motion-compensated transform coding. Section 3 and 4 discuss respectively uncorrelated and correlated displacement errors, and provide numerical results.

2. MODEL FOR MOTION-COMPENSATED TRANSFORM CODING

Let $\mathbf{v}[l]$ and $\mathbf{c}_k[l]$ 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. We interpret $\mathbf{c}_k$ as the k-th of K motion-compensated pictures to be coded.

Obviously, motion compensation should work best if we compensate the true displacement of the scene exactly. Less accurate compensation will degrade the performance. To capture the limited accuracy of motion compensation, we associate a vector-valued displacement error $\mathbf{\Delta}_k$ with the k-th motion-compensated picture $\mathbf{c}_k$. 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 motion-compensated pictures are shifted versions of the "clean" video signal $\mathbf{v}$ and distorted by independent additive white Gaussian noise $\mathbf{n}_k$. The shift is determined by the vector-valued displacement error $\mathbf{\Delta}_k$ of the k-th motion-compensated picture. For that, the ideal reconstruction of the band-limited signal $\mathbf{v}[l]$ is shifted by the continuous valued displacement error and re-sampled on the original orthogonal grid. The noise signals $\mathbf{n}_\mu$ and $\mathbf{n}_\nu$ are mutually statistically independent for $\mu \neq \nu$.

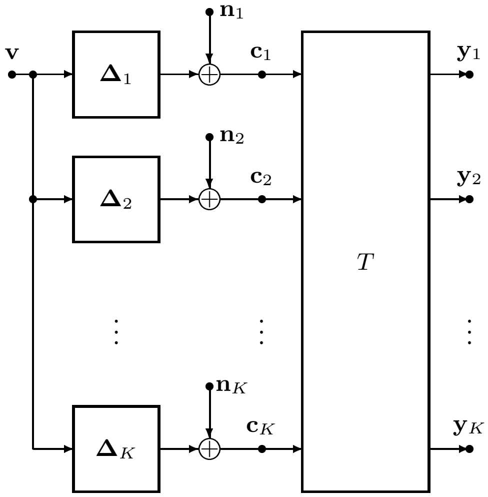

Fig. 1. Model for a group of K motion-compensated pictures.

Fig. 1 depicts the model for K motion-compensated pictures $\mathbf{c}_k[l]$ which are jointly transformed by T with K output signals $\mathbf{y}_k[l]$. We assume that T is linear, unitary, and decorrelating.

Assume that $\mathbf{v}$ and $\mathbf{c}_k$ are generated by a jointly wide-sense stationary random process with the real-valued scalar two-dimensional power spectral densities $\Phi_{\mathbf{vv}}(\omega)$ and $\Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega)$. The power spectral densities $\Phi_{\mathbf{c}_\mu\mathbf{c}_\nu}(\omega)$ are elements in the power spectral density matrix of the motion-compensated pictures $\Phi_{\mathbf{cc}}$. The power spectral density matrix of the decorrelated signal $\Phi_{\mathbf{yy}}$ is given by $\Phi_{\mathbf{cc}}$ and the transform T,

$$\Phi_{\mathbf{yy}}(\omega) = T(\omega)\Phi_{\mathbf{cc}}(\omega)T^H(\omega), \qquad (1)$$

where T^H denotes the Hermitian of T and $\omega = (\omega_x, \omega_y)^T$ the vector-valued frequency.

We adopt the expressions for the cross spectral densities $\Phi_{\mathbf{c}_\mu \mathbf{c}_\nu}$ from [2], where the displacement errors $\mathbf{\Delta}_k$ are interpreted as random variables which are statistically independent from $\mathbf{v}$:

$$\Phi_{\mathbf{c}_\mu \mathbf{c}_\nu}(\omega) = E\left\{ e^{-j\omega^T(\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)} \right\} \Phi_{\mathbf{vv}}(\omega) + \Phi_{\mathbf{n}_\mu \mathbf{n}_\nu}(\omega) \quad (2)$$

Like in [2], we assume a power spectrum $\Phi_{\mathbf{vv}}$ that corresponds to an exponentially decaying isotropic autocorrelation function with a correlation coefficient $\rho_{\mathbf{v}} = 0.93$.

For the k-th displacement error $\mathbf{\Delta}_k$, a 2-D 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 K motion-compensated pictures. This is reasonable because all pictures are compensated with the same accuracy. Further, the pairs $(\mathbf{\Delta}_\mu, \mathbf{\Delta}_\nu)$ are assumed to be jointly Gaussian random variables [3]. For simplicity, we assume that 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 motion-compensated pictures. The above assumptions are summarized by the covariance matrix of a displacement error component.

$$C_{\mathbf{\Delta}_x \mathbf{\Delta}_x} = \sigma^2_{\mathbf{\Delta}} \begin{pmatrix} 1 & \rho_{\mathbf{\Delta}} & \cdots & \rho_{\mathbf{\Delta}} \\ \rho_{\mathbf{\Delta}} & 1 & \cdots & \rho_{\mathbf{\Delta}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{\mathbf{\Delta}} & \rho_{\mathbf{\Delta}} & \cdots & 1 \end{pmatrix}. \quad (3)$$

Since the covariance matrix is non-negative definite [4], the correlation coefficient $\rho_{\mathbf{\Delta}}$ in (3) has the limited range

$$\frac{1}{1-K} \le \rho_{\mathbf{\Delta}} \le 1 \quad \text{for} \quad K = 2, 3, 4, \ldots, \quad (4)$$

which is dependent on the number of motion-compensated pictures K.

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

$$E\left\{ e^{-j\omega^T \mathbf{\Delta}_k} \right\} := P(\omega, \sigma^2_{\mathbf{\Delta}}) = e^{-\frac{1}{2}\omega^T \omega \sigma^2_{\mathbf{\Delta}}} \quad (5)$$

The expected value 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 (2)

$$E\left\{ e^{-j\omega^T(\mathbf{\Delta}_\mu - \mathbf{\Delta}_\nu)} \right\} = P\left(\omega, 2\sigma^2_{\mathbf{\Delta}}(1 - \rho_{\mathbf{\Delta}})\right) \quad \text{for} \quad \mu \ne \nu. \quad (6)$$

For $\mu = \nu$, the expected value in (2) is equal to one. With that, we obtain for the power spectral density matrix of the motion-compensated pictures:

$$\frac{\Phi_{\mathbf{cc}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} = \begin{pmatrix} 1+\alpha(\omega) & P(\omega,\sigma^2) & \cdots & P(\omega,\sigma^2) \\ P(\omega,\sigma^2) & 1+\alpha(\omega) & \cdots & P(\omega,\sigma^2) \\ \vdots & \vdots & \ddots & \vdots \\ P(\omega,\sigma^2) & P(\omega,\sigma^2) & \cdots & 1+\alpha(\omega) \end{pmatrix} \quad (7)$$

$\alpha(\omega)$ is the normalized spectral density of the noise $\Phi_{\mathbf{n}_k \mathbf{n}_k}(\omega)$ with respect to the spectral density of the "clean" video signal.

$$\alpha(\omega) = \frac{\Phi_{\mathbf{n}_k \mathbf{n}_k}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} \quad \text{for} \quad k = 1, 2, \ldots, K \quad (8)$$

For T, we assume an energy preserving and decorrelating transform: the KLT. The eigenvalues of the power spectral density matrix $\Phi_{\mathbf{cc}}$ are $\lambda_1 = [1 + \alpha(\omega) + (K-1)P(\omega,\sigma^2)]\Phi_{\mathbf{vv}}$ and $\lambda_{2,3,\ldots,K} = [1 + \alpha(\omega) - P(\omega,\sigma^2)]\Phi_{\mathbf{vv}}$. The power spectral density matrix of the transformed signals is diagonal.

$$\frac{\Phi_{\mathbf{yy}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} = \begin{pmatrix} 1+\alpha+(K-1)P & 0 & \cdots & 0 \\ 0 & 1+\alpha-P & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1+\alpha-P \end{pmatrix} \quad (9)$$

The first eigenvector just adds all components and scales with $1/\sqrt{K}$. For the remaining eigenvectors, any orthonormal basis can be used that is orthogonal to the first eigenvector. That is, the KLT for our signal model is not dependent on ω.

The rate difference [5], [2] is used to measure the improved compression efficiency for each motion-compensated picture k.

$$\Delta R_k = \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(\frac{\Phi_{\mathbf{y}_k \mathbf{y}_k}(\omega)}{\Phi_{\mathbf{c}_k \mathbf{c}_k}(\omega)} \right) d\omega \quad (10)$$

It represents the maximum bit-rate reduction (in bit/sample) possible by optimum encoding of the transformed signal $\mathbf{y}_k$, compared to optimum intra-frame encoding of the signal $\mathbf{c}_k$ for Gaussian wide-sense stationary signals for the same mean squared reconstruction error. A negative ΔR_k 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 overall rate difference ΔR is the average over all motion-compensated pictures and is used to evaluate the efficiency of motion-compensated transform coding.

$$\Delta R = \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{2K} \log_2 \left(\frac{\prod_{k=1}^{K} \Phi_{\mathbf{y}_k \mathbf{y}_k}(\omega)}{\prod_{k=1}^{K} \Phi_{\mathbf{c}_k \mathbf{c}_k}(\omega)} \right) d\omega \quad (11)$$

Assuming the KLT, we obtain for the overall rate difference with (9)

$$\Delta R = \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{K-1}{2K} \log_2 \left(1 - \frac{P(\omega,\sigma^2)}{1+\alpha(\omega)} \right) + \frac{1}{2K} \log_2 \left(1 + (K-1)\frac{P(\omega,\sigma^2)}{1+\alpha(\omega)} \right) d\omega. \quad (12)$$

The case of a very large number of motion-compensated pictures is of special interest for the comparison to DPCM with motion-compensated prediction.

$$\Delta R_{K\to\infty} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(1 - \frac{P(\omega,\sigma^2)}{1+\alpha(\omega)} \right) d\omega \quad (13)$$

According to [1], the performance of motion-compensated prediction with optimum Wiener filter achieves a rate difference of

$$\Delta R_{\mathrm{MCP}} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(1 - \frac{P^2(\omega,\sigma^2_{\mathbf{\Delta}})}{[1+\alpha(\omega)]^2} \right) d\omega. \quad (14)$$

Assuming uncorrelated displacement errors ($\sigma^2 = 2\sigma^2_{\mathbf{\Delta}}$), the performance of motion-compensated transform coding and MCP differs only in the influence of the residual noise power spectrum $\alpha(\omega)$. That is, assuming no residual noise, both approaches demonstrate identical performance.

3. UNCORRELATED DISPLACEMENT ERRORS

In the following, we investigate motion-compensated transform coding and provide numerical results for comparison.

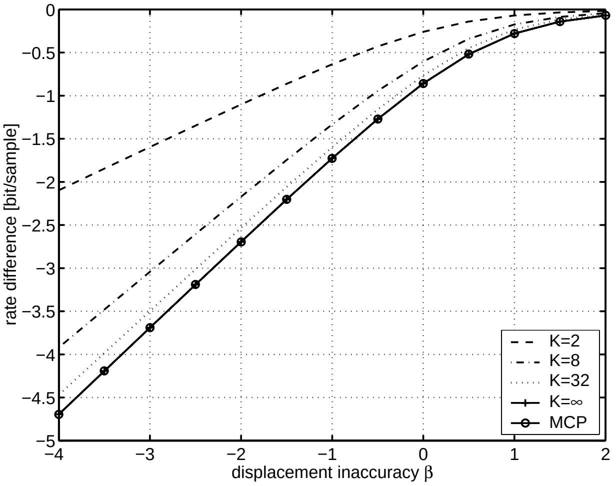

Fig. 2. Rate difference to intra-frame coding vs. displacement inaccuracy for a group of K pictures. The displacement errors are uncorrelated and the residual noise is -100 dB. For reference, the performance of motion-compensated prediction with optimum Wiener filter is also plotted.

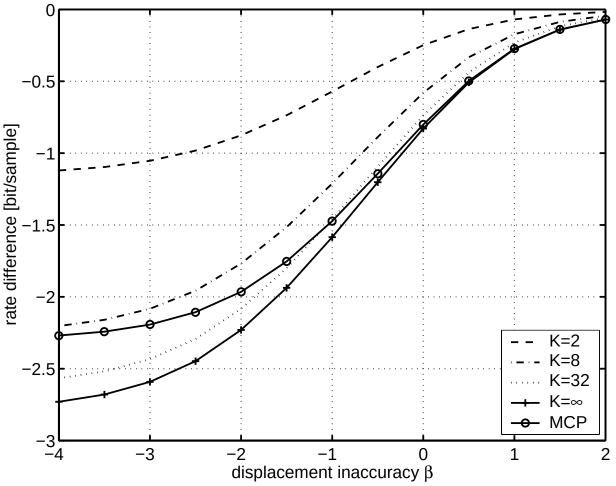

Fig. 3. Rate difference to intra-frame coding vs. displacement inaccuracy for a group of K pictures. The displacement errors are uncorrelated and the residual noise is -30 dB. For reference, the performance of motion-compensated prediction with optimum Wiener filter is also plotted.

Figs. 2 and 3 depict the overall rate difference over the displacement inaccuracy for motion-compensated transform coding with 2, 8, and 32 pictures. The residual noise RNL is -100 dB and -30 dB, respectively. A residual noise level of -100 dB can be regarded as the noiseless case, whereas -30 dB reflects the noise level of compressed video. The limit of a very large number of pictures and the performance of motion-compensated prediction according to (14) is also plotted. 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 [2]. The displacement error variance is

$$\sigma^2_{\mathbf{\Delta}} = \frac{2^{2\beta}}{12}. \tag{15}$$

We investigate the slope of the rate difference (13) in the noiseless case $\alpha \to 0$ for $K \to \infty$ by applying a Taylor series expansion of first order for the function P.

$$1 - \frac{P(\omega, \sigma^2)}{1+\alpha(\omega)} \approx \sigma^2_{\mathbf{\Delta}} \omega^T \omega (1 - \rho_\Delta) \quad \text{for} \quad \sigma^2_{\mathbf{\Delta}} \to 0, \rho_\Delta \neq 1 \tag{16}$$

When inserting this in (13), we obtain a slope of 1 bit per inaccuracy step for motion-compensated transform coding.

$$\Delta R_{K\to\infty} \approx \beta + const. \quad \text{for} \quad \sigma^2_{\mathbf{\Delta}} \to 0, \rho_\Delta \neq 1 \tag{17}$$

With this assumptions, motion-compensated transform coding and motion-compensated prediction show the same behavior in the noiseless case. This can also be observed in Fig. 2, where the residual noise is negligible.

If residual noise is present, motion-compensated transform coding outperforms motion-compensated prediction by at most 0.5 bit/sample for very accurate motion compensation. This can be observed in Fig. 3 and by comparing (13) with (14).

4. CORRELATED DISPLACEMENT ERRORS

So far, we assumed uncorrelated displacement errors for motion compensation. In the following, we investigate the influence of the displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$ on the performance of motion-compensated transform coding.

Fig. 4 depicts rate difference over displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$ for a group of 2, 8, and 32 pictures at a residual noise level of -50 dB. The limit of a very large number of pictures and the performance of motion-compensated prediction according to (14) is also plotted. We assume that for motion-compensated prediction the displacement errors are uncorrelated. We observe that for increasing correlation coefficient the performance of motion-compensated transform coding improves and outperforms motion-compensated prediction for a large number of pictures. Maximally negatively correlated displacement errors cause an inferior performance.

Figs. 5 and 6 capture rate difference over displacement inaccuracy for a group of 2, 8, and 32 pictures at a residual noise level of -30 dB. The displacement error correlation coefficient $\rho_{\mathbf{\Delta}}$ is 0.5 and 1, respectively. The limit of a very large number of pictures and the performance of motion-compensated prediction according to (14) is also plotted. Comparing Fig. 3, 5, and 6 with $\rho_{\mathbf{\Delta}} = 0$, $\rho_{\mathbf{\Delta}} = 0.5$, and $\rho_{\mathbf{\Delta}} = 1$, respectively, we observe that the influence of motion accuracy on the rate difference is reduced for

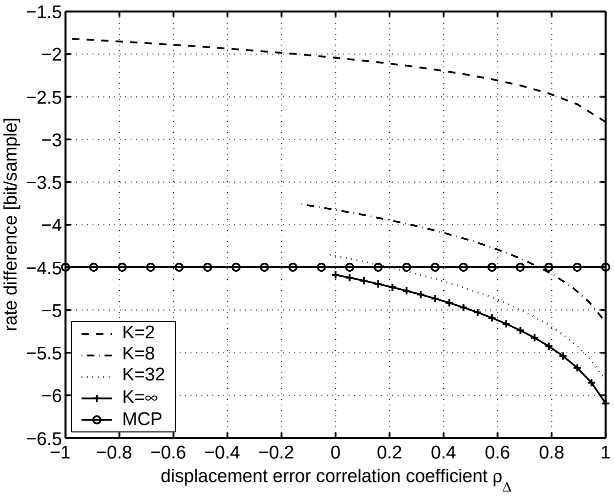

Fig. 4. Rate difference to intra-frame coding vs. displacement error correlation coefficient for a group of K pictures. Very accurate motion compensation is assumed ($\beta = -4$) and the residual noise is -50 dB. For reference, the performance of motion-compensated prediction with optimum Wiener filter is also plotted.

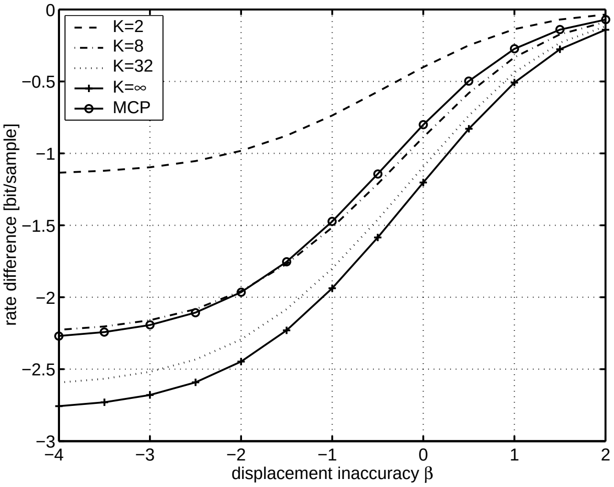

Fig. 5. Rate difference to intra-frame coding vs. displacement inaccuracy for a group of K pictures. The displacement errors are correlated with $\rho_\Delta = 0.5$ and the residual noise is -30 dB. For reference, the performance of motion-compensated prediction with optimum Wiener filter is also plotted.

increasing correlation coefficient. For maximally positively correlated displacement errors, the efficiency of motion-compensated transform coding depends only on the residual noise level.

5. CONCLUSIONS

We presented an efficiency analysis of motion-compensated transform coding. For that, we assume that K pictures are motion-compensated up to a displacement error with given variance and distorted by statistically independent additive white Gaussian noise.

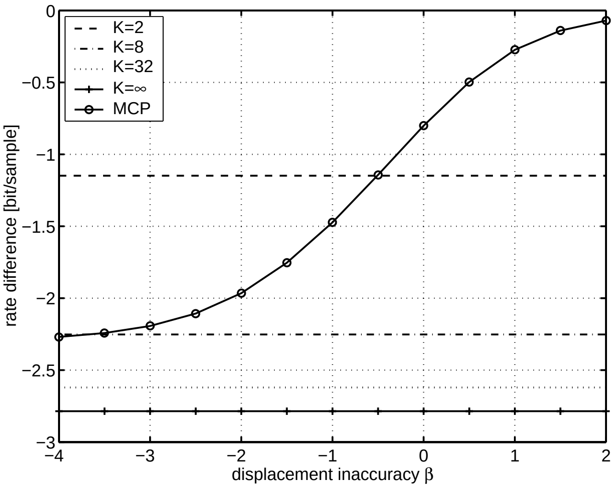

Fig. 6. Rate difference to intra-frame coding vs. displacement inaccuracy for a group of K pictures. The displacement errors are maximally correlated ($\rho_\Delta = 1$) and the residual noise is -30 dB. For reference, the performance of motion-compensated prediction with optimum Wiener filter is also plotted.

Further, we decorrelate the K motion-compensated pictures by the KLT and determine the maximum bit-rate reduction possible by optimum encoding of the transformed signal, compared to optimum intra-frame encoding of the motion-compensated signal. We show that the performance for uncorrelated displacement errors in the noiseless case is identical to motion-compensated prediction. In the presence of residual noise, the presented scheme demonstrates an improvement of at most 0.5 bit/sample. In the case of correlated displacement errors, we pointed out that compression efficiency improves for positively correlated displacement errors and that the performance is only limited by the residual noise.

6. ACKNOWLEDGMENT

The authors would like to thank Prashant Ramanathan for helpful discussions.

7. REFERENCES

[1] B. Girod, “The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences,” *IEEE Journal on Selected Areas in Communications*, vol. SAC-5, no. 7, pp. 1140–1154, Aug. 1987.

[2] 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.

[3] M. Flierl and B. Girod, “Multihypothesis Motion Estimation for Video Coding,” in *Proceedings of the Data Compression Conference*, Snowbird, Utah, Mar. 2001, pp. 341–350.

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

[5] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice-Hall, Englewood Cliffs, 1971.