DISTRIBUTED CODING OF DYNAMIC SCENES

Markus Flierl

Signal Processing Institute
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
markus.flierl@epfl.ch

ABSTRACT

We address the problem of compressing correlated distributed video signals that are captured from a dynamic scene. The correlated video signals originate from cameras that are not co-located or that cannot cooperate to directly exploit their correlation. However, the decoder is able to exploit the coded information from all cameras to achieve the best reconstruction of the correlated video signals. Our distributed coding scheme is based on a motion-compensated lifted wavelet transform to exploit the temporal correlation of the camera signals. The correlation among the video signals is considered by coset-encoding the quantized wavelet transform coefficients. The experimental results demonstrate that conditional decoding can reduce the bit-rate of one sequence by up to 20% when compared to independent decoding. Further, we consider theoretically the associated rate-distortion problem with side information and determine the optimal conditional Karhunen-Loeve transform for video coding with side information and outline the performance bounds.

1. INTRODUCTION

Scene information that is acquired by more than one sensor can be coded efficiently if the correlation among sensor signals is exploited. In one possible compression scenario, encoders of the sensor signals are connected and compress the sensor signals jointly. In an alternative compression scenario, each encoder operates independently but relies on a joint decoding unit that receives all coded sensor signals. This is also known as distributed source coding. A special case of this scenario is source coding with side information. Wyner and Ziv showed that for certain cases, the encoder does not need the side information to which the decoder has access to achieve the rate-distortion bound [1]. A practical algorithm for source coding with side information using syndromes is suggested by Pradhan and Ramchandran [2]. For transform-based source coding, Vetterli et al. studied the distributed and conditional Karhunen-Loeve transform [3]. Examples of applied research on distributed source coding are enhancing analog image transmission systems using digital side information, Wyner-Ziv coding of inter-pictures in video sequences, and distributed compression of light field images. This paper discusses a distributed source coding scenario where the sensors are video cameras that capture a dynamic scene. The video signals are encoded with a motion-compensated lifted wavelet transform which approximates the temporal Karhunen-Loeve transform for video signals [4, 5]. The distributed video coding scheme employs coset-encoding and considers the video side information at the decoder.

The paper is organized as follows: Section 2 outlines our distributed coding scheme for dynamic scenes. We discuss the used motion-compensated wavelet transform as well as the coset-encoding of the quantized transform coefficients. We conclude with experimental results. Section 3 explores the efficiency of video coding with side information. With a model for transform-coded video signals, we determine the temporal conditional Karhunen-Loeve transform and discuss bounds for the coding gain due to side information.

2. DISTRIBUTED CODING SCHEME

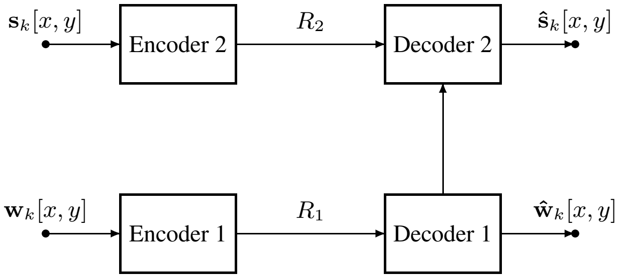

Fig. 1. Distributed coding scheme for dynamic scenes.

Fig. 1 depicts the distributed coding scheme for dynamic scenes. The dynamic scene is represented by the image sequences $\mathbf{s}_k[x,y]$ and $\mathbf{w}_k[x,y]$. The coding scheme comprises of *Encoder 1* and *Encoder 2* that operate independently as well as of *Decoder 2* that is dependent on results of *Decoder 1*.

2.1. Motion-Compensated Transform

Each encoder in Fig. 1 exploits the correlation between successive pictures by employing a motion-compensated temporal transform for groups of K pictures (GOP). We perform a dyadic decomposition with a motion-compensated Haar wavelet as depicted in Fig. 2. The temporal transform provides K output pictures that are decomposed by a spatial 8×8 DCT. The motion information that is required for the motion-compensated wavelet transform is estimated in each decomposition level depending on the results of the lower level. The correlation of motion information between two image sequences is not exploited yet. Fig. 2 shows the Haar wavelet with motion-compensated lifting steps. The even frames of the video sequence $\mathbf{s}_{2k}$ are used to predict the odd frames $\mathbf{s}_{2k+1}$ with the estimated motion vector $\hat{d}_{2k,2k+1}$. The prediction step is followed by an update step which uses the negative motion vector as an approximation. We use a block-size of 16×16 and half-pel accurate motion compensation with bi-linear interpolation in the prediction step and select the motion vectors such that they minimize a Lagrangian cost function based on the squared error in the high-band $\mathbf{h}_k$. Additional scaling factors in low- and high-band are necessary to normalize the transform.

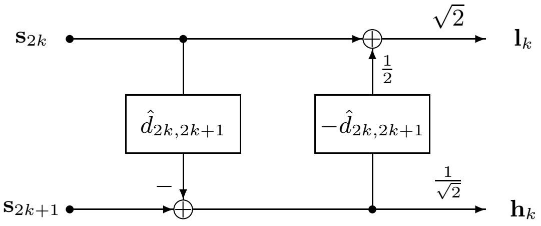

Fig. 2. Haar wavelet with motion-compensated lifting steps.

Encoder 1 in Fig. 1 encodes the side information for *Decoder 2* and does not employ distributed source coding principles yet. It uses scalar quantizers to represent the DCT coefficients of all temporal bands. The quantized coefficients are simply run-length encoded. On the other hand, *Encoder 2* is designed for distributed source coding and employs coset-encoding of the quantized DCT coefficients for all temporal bands.

2.2. Coset-Encoding of Quantized Transform Coefficients

The 8×8 DCT coefficients of *Encoder 2* are encoded with the DISCUS framework [2]. Currently, we use uniform scalar quantization and construct the cosets in a memoryless fashion. Fig. 3 explains this coset-coding scheme.

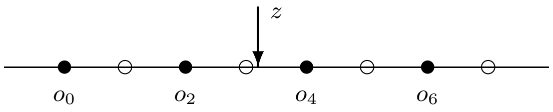

Fig. 3. Coset-coding of quantized transform coefficients. Assume that *Encoder 2* transmits with a rate R_2 of 1 bit per transform coefficient and utilizes two cosets $\mathcal{C}_0 = \{o_0, o_2, o_4, o_6\}$ and $\mathcal{C}_1 = \{o_1, o_3, o_5, o_7\}$ for encoding. Now, the transform coefficient o_4 shall be encoded and the encoder sends one bit to signal coset $\mathcal{C}_0$. With the help of the side information coefficient z, the decoder is able to decode o_4 correctly. If *Encoder 2* does not send any bit, the decoder will decode o_3 and we observe a decoding error.

Decoder 2 receives coset indices from *Encoder 2*. The binary representation of these indices reflects the nested construction of the cosets. If a coset index is represented by i bits, the indicated coset out of 2^i cosets is used for decoding with side information. If the number of cosets doubles, the Euclidean distance between the representatives in one coset also doubles. This guarantees reliable decoding if sufficient bits are received. *Encoder 2* receives from *Decoder 2* side information for optimal conditional quantization. With that, *Decoder 2* can decode the coefficient without error.

2.3. Experimental Results

For the experiments, we selected the stereoscopic MPEG-4 sequences *Funfair* and *Tunnel* in QCIF resolution. We divided each view with 224 frames at 30 fps into groups of $K = 32$ pictures. The GOPs of the left view are encoded with *Encoder 1* at high quality by setting the quantization parameter $QP = 2$. This coded version of the left view is used as side information for *Decoder 2* to decode the right view.

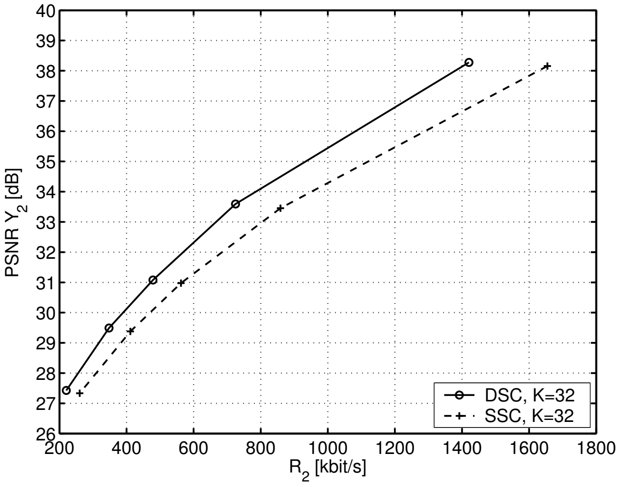

Fig. 4. Luminance PSNR vs. total bit-rate of the distributed codec DSC for the sequence *Funfair 2* (right view). The reference coding scheme SSC does not utilize the side information for decoding.

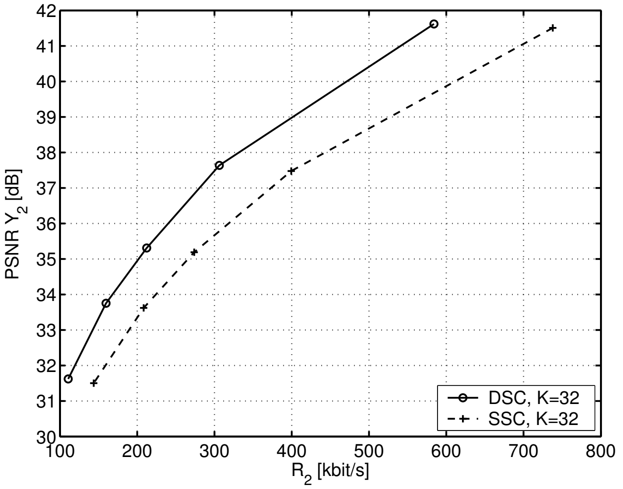

Fig. 5. Luminance PSNR vs. total bit-rate of the distributed codec DSC for the sequence *Tunnel 2* (right view). The reference coding scheme SSC does not utilize the side information for decoding.

Figs. 4 and 5 show the luminance PSNR over the total bit-rate of the distributed codec *Encoder 2* for the sequences *Funfair 2* and *Tunnel 2*, respectively. These sequences are the right views of the stereoscopic sequences. The rate-distortion points are obtained by setting different quantization parameters for the scalar quantizer in *Encoder 2*. We observe that the use of side information at the decoder reduces the bit-rate of *Funfair 2* by up to 15% (which corresponds to a gain of up to 1.5 dB). *Tunnel 2* is stronger correlated to the side information *Tunnel 1* and achieves bit-rate savings of up to 20% (which corresponds to a gain of up to 2 dB).

3. EFFICIENCY OF VIDEO CODING WITH SIDE INFORMATION

In the following, we outline a mathematical model to study video coding with side information in more detail. We derive performance bounds and compare to coding without side information.

3.1. Model for Transform-Coded Video Signals

We build upon a model for motion-compensated subband coding of video that is outlined in [4, 5]. Let $\mathbf{s}_k = \{\mathbf{s}_k[x,y], (x,y) \in \Pi\}$ be scalar random fields over a two-dimensional orthogonal grid Π with horizontal and vertical spacing of 1. In Fig. 6, we assume that the pictures $\mathbf{s}_k$ are shifted versions of the model picture $\mathbf{v}$ and degraded by independent additive white Gaussian noise $\mathbf{n}_k$ [6]. $\boldsymbol{\Delta}_k$ is the displacement error in the k-th picture, statistically independent from the model picture $\mathbf{v}$ and the noise $\mathbf{n}_k$ but correlated to other displacement errors. We assume a 2-D normal distribution with variance $\sigma^2_{\boldsymbol{\Delta}}$ and zero mean where the x- and y-components are statistically independent.

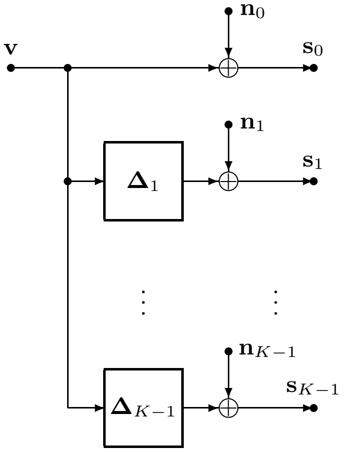

Fig. 6. Signal model for a group of K pictures.

From [4, 5], we adopt the matrix of the power spectral densities of the pictures $\mathbf{s}_k$ and normalize it with respect to the power spectral density of the model picture $\mathbf{v}$. We write it also with the identity matrix I and the matrix $\mathbf{1}\mathbf{1}^T$ with all entries equal to 1.

$$
\begin{aligned}
\frac{\Phi_{\mathbf{ss}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} &= \begin{pmatrix} 1+\alpha(\omega) & P(\omega) & \cdots & P(\omega) \\ P(\omega) & 1+\alpha(\omega) & \cdots & P(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ P(\omega) & P(\omega) & \cdots & 1+\alpha(\omega) \end{pmatrix} \\
&= [1+\alpha(\omega) - P(\omega)]\, I + P(\omega)\mathbf{1}\mathbf{1}^T \qquad (1)
\end{aligned}
$$

$\alpha = \alpha(\omega)$ is the normalized power spectral density of the noise $\Phi_{\mathbf{n}_k\mathbf{n}_k}(\omega)$ with respect to the model picture $\mathbf{v}$.

$$\alpha(\omega) = \frac{\Phi_{\mathbf{n}_k\mathbf{n}_k}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} \quad \text{for} \quad k = 0, 1, \ldots, K-1 \qquad (2)$$

$P = P(\omega)$ is the characteristic function of the continuous 2-D Gaussian displacement error.

$$P(\omega) = E\left\{ e^{-j\omega^T \boldsymbol{\Delta}_k} \right\} = e^{-\frac{1}{2}\omega^T\omega\sigma^2_{\boldsymbol{\Delta}}} \qquad (3)$$

3.2. Rate-Distortion with Video Side Information

Now, we consider the distributed coding scheme in Fig. 1 at high rates such that the reconstructed side information approaches the original side information $\hat{\mathbf{w}}_k \to \mathbf{w}_k$. With that, we have a Wyner-Ziv scheme (Fig. 7) and the rate-distortion function R^* of *Encoder 2* is bounded by the conditional rate-distortion function [1].

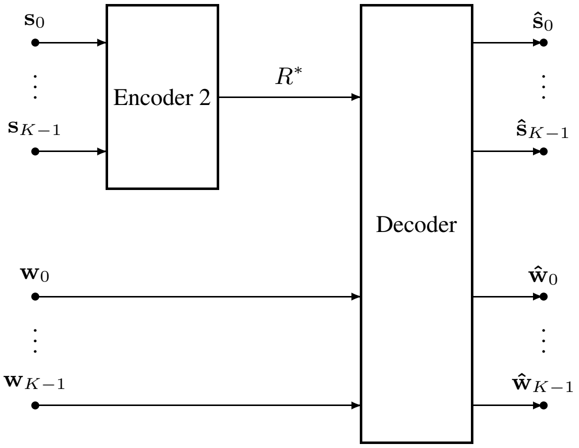

Fig. 7. Coding of K pictures $\mathbf{s}_k$ at rate R^* with side information of K pictures $\mathbf{w}_k$ at the decoder.

We model the side information as a noisy version of the video signal to be encoded, i.e. $\mathbf{w}_k = \mathbf{s}_k + \mathbf{u}_k$, and assume that the noise $\mathbf{u}_k$ is also Gaussian with variance $\sigma^2_{\mathbf{u}}$ and independent of $\mathbf{s}_k$. In this case, the matrix of the power spectral densities of the side information pictures is simply $\Phi_{\mathbf{ww}}(\omega) = \Phi_{\mathbf{ss}}(\omega) + \Phi_{\mathbf{uu}}(\omega)$ with the matrix of the power spectral densities of the side information noise $\Phi_{\mathbf{uu}}(\omega) = \gamma(\omega)\Phi_{\mathbf{vv}}(\omega)I$. $\gamma = \gamma(\omega)$ is the normalized power spectral density of the side information noise $\Phi_{\mathbf{u}_k\mathbf{u}_k}(\omega)$ with respect to the model picture $\mathbf{v}$.

$$\gamma(\omega) = \frac{\Phi_{\mathbf{u}_k\mathbf{u}_k}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} \quad \text{for} \quad k = 0, 1, \ldots, K-1 \qquad (4)$$

With these assumptions, the rate-distortion function R^* of *Encoder 2* is equal to the conditional rate-distortion function [1]. Now, it is sufficient to use the conditional Karhunen-Loeve transform [3] to code video signals with side information and achieve the conditional rate-distortion function.

3.3. Conditional Karhunen-Loeve Transform

In the case of motion-compensated transform coding of video with side information, the conditional Karhunen-Loeve transform is required to obtain the performance bounds. We determine the well known conditional power spectral density matrix $\Phi_{\mathbf{s}|\mathbf{w}}(\omega)$ of the video signal $\mathbf{s}_k$ given the video side information $\mathbf{w}_k$.

$$\Phi_{\mathbf{s}|\mathbf{w}}(\omega) = \Phi_{\mathbf{ss}}(\omega) - \Phi^H_{\mathbf{ws}}(\omega)\Phi^{-1}_{\mathbf{ww}}(\omega)\Phi_{\mathbf{ws}}(\omega) \qquad (5)$$

With the model in Section 3.1 and the assumptions in Section 3.2, we obtain for the normalized conditional spectral density matrix

$$
\begin{aligned}
\frac{\Phi_{\mathbf{s}|\mathbf{w}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} &= \frac{1+\alpha-P}{1+\alpha+\gamma-P}\gamma I + \\
&\frac{P}{1+\alpha+\gamma-P} \cdot \frac{\gamma}{1+\alpha+\gamma+[K-1]P}\gamma\mathbf{1}\mathbf{1}^T. \qquad (6)
\end{aligned}
$$

For our signal model, the conditional KLT is as follows: 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. The Haar wavelet that we use for our coding scheme meets these requirements. Finally, K eigendensities are needed to determine the performance bounds:

$$\begin{aligned} \frac{\Lambda_0^*(\omega)}{\Phi_{\mathbf{vv}}(\omega)} &= \frac{1+\alpha+\frac{\gamma K P}{1+\alpha+\gamma+[K-1]P} - P}{1+\alpha+\gamma-P}\gamma \\ \frac{\Lambda_k^*(\omega)}{\Phi_{\mathbf{vv}}(\omega)} &= \frac{1+\alpha-P}{1+\alpha+\gamma-P}\gamma \quad k=1,2,\dots,K-1 \quad (7) \end{aligned}$$

3.4. Coding Gain due to Side Information

With the conditional eigendensities, we are able to determine the coding gain due to side information. We normalize the conditional eigendensities $\Lambda_k^*(\omega)$ with respect to the eigendensities $\Lambda_k(\omega)$ that we obtain for coding without side information as $\Lambda_k^*(\omega) \to \Lambda_k(\omega)$ for $\gamma(\omega) \to \infty$.

$$\begin{aligned} \frac{\Lambda_0^*(\omega)}{\Lambda_0(\omega)} &= \frac{\gamma}{1+\alpha+\gamma-P} \cdot \frac{1+\alpha+\frac{\gamma K P}{1+\alpha+\gamma+[K-1]P} - P}{1+\alpha+[K-1]P} \\ \frac{\Lambda_k^*(\omega)}{\Lambda_k(\omega)} &= \frac{\gamma}{1+\alpha+\gamma-P} \quad k=1,2,\dots,K-1 \qquad (8) \end{aligned}$$

The rate difference is used to measure the improved compression efficiency for each picture k in the presence of side information.

$$\Delta R_k^* = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(\frac{\Lambda_k^*(\omega)}{\Lambda_k(\omega)} \right) d\omega \qquad (9)$$

It represents the maximum bit-rate reduction (in bit/sample) possible by optimum encoding of the eigensignal with side information, compared to optimum encoding of the eigensignal without side information for Gaussian wide-sense stationary signals for the same mean square reconstruction error. The overall rate difference ΔR^* is the average over all K eigensignals.

Figs. 8 and 9 depict the overall rate difference for a residual noise level RNL $= 10\log_{10}(\sigma_{\mathbf{n}}^2)$ of -30 dB over the c-SNR $= 10\log_{10}([\sigma_{\mathbf{v}}^2 + \sigma_{\mathbf{n}}^2]/\sigma_{\mathbf{u}}^2)$ and the displacement inaccuracy $\beta = \log_2(\sqrt{12}\sigma_{\mathbf{\Delta}})$, respectively. Note that the variance of the model picture $\mathbf{v}$ is normalized to $\sigma_{\mathbf{v}}^2 = 1$. We observe that side information is more beneficial if temporal correlation remains due to small GOP sizes K. For highly correlated video signals, the gain due to side information increases by 1 bit/sample if the c-SNR increases by 6 dB. For $K = 32$, half-pel accurate motion compensation ($\beta = -1$), and a c-SNR of 20 dB, the rate difference is limited to -0.3 bit/sample which corresponds to a gain of 1.8 dB.

4. CONCLUSIONS

This paper discusses the problem of compressing correlated distributed video signals that are captured from a dynamic scene. We consider theoretically the associated rate-distortion problem with side information and determine the optimal conditional KLT.

5. REFERENCES

[1] A.D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, Jan. 1976.

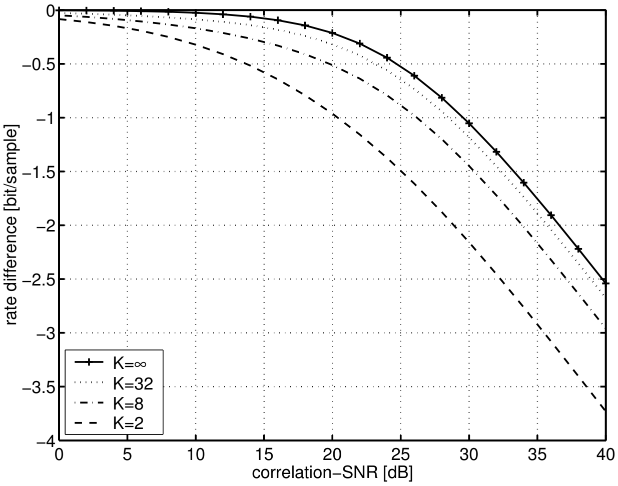

Fig. 8. Rate difference to motion-compensated transform coding without side information vs. correlation-SNR for groups of K pictures. The displacement inaccuracy β is -1 (half-pel accuracy) and the residual noise is -30 dB.

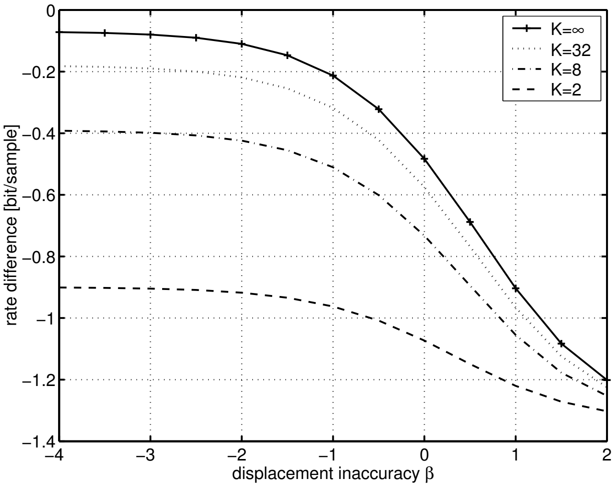

Fig. 9. Rate difference to motion-compensated transform coding without side information vs. displacement inaccuracy β for groups of K pictures. The residual noise is -30 dB and the correlation-SNR is 20 dB.

[2] S.S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[3] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loève transforms," in *Proceedings of the Data Compression Conference*, Snowbird, UT, Mar. 2003, pp. 283–292.

[4] M. Flierl and B. Girod, "Investigation of motion-compensated lifted wavelet transforms," in *Proceedings of the Picture Coding Symposium*, Saint-Malo, France, Apr. 2003, pp. 59–62.

[5] M. Flierl and B. Girod, *Video Coding with Superimposed Motion-Compensated Signals: Applications to H.264 and Beyond*, Kluwer Academic Publishers, 2004.

[6] 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.