Distributed Coding of Dynamic Scenes with Motion-Compensated Wavelets

Markus Flierl and Pierre Vandergheynst
Signal Processing Institute
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
{markus.flierl,pierre.vandergheynst}@epfl.ch

***Abstract*— This paper discusses distributed coding of two correlated video signals. The video signals are captured from a dynamic scene where each signal is temporally decorrelated by a motion-compensated Haar wavelet. The two cameras operate independently, however, the central decoder is able to exploit the coded information from all cameras to achieve the best reconstruction of the correlated video signals. The coding system utilizes nested lattice codes for the transform coefficients and exploits side information at the decoder. The efficiency of the decoder is improved by disparity compensation of one video signal. When compared to decoding without side information, decoding of the quantized transform coefficients with side information reduces the bit-rate of our test sequence by up to 5%. Further, we observe bit-rate savings of up to 8% with disparity compensation at the decoder and decoding of transform coefficients with side information.**

I. 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].

Examples of applied research on distributed source coding are enhancing analog image transmission systems using digital side information [2], Wyner-Ziv coding of inter-pictures in video sequences [3], and distributed compression of light field images [4]. 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 motion-compensated temporal Karhunen-Loeve transform for video signals [5]. The distributed video coding scheme employs nested lattice codes and considers disparity-compensated video side information at the decoder.

The paper is organized as follows: Section II outlines our distributed coding scheme for dynamic scenes. We discuss the used motion-compensated temporal transform, the coset-encoding of transform coefficients with nested lattice codes, decoding with side information, and enhancing the side information by disparity compensation. Section III provides experimental rate-distortion results for decoding of video signals with side information. Moreover, it discusses the relation between the level of temporal decorrelation and the efficiency of decoding with side information.

II. Distributed Coding Scheme

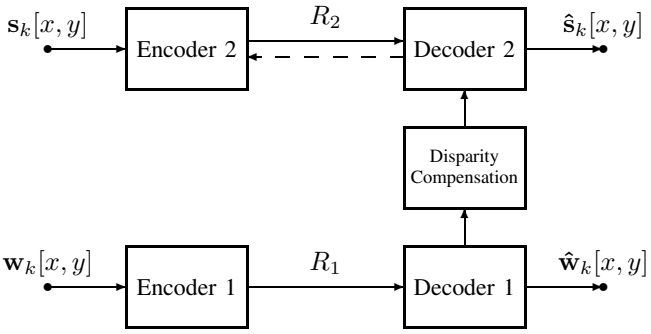

Fig. 1. Distributed coding scheme with disparity compensation.

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 *Decoder 1*. The side information for *Decoder 2* can be improved by considering the spatial camera positions and performing disparity compensation. As the video signals are not stationary, *Decoder 2* is decoding with feed-back.

A. Motion-Compensated Temporal 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, i.e. coded motion

vectors are not part of the side information. 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.

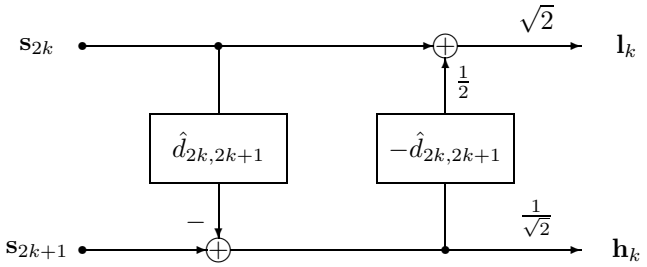

Fig. 2. Haar wavelet with motion-compensated lifting steps.

Encoder 1 in Fig. 1 encodes the side information for *Decoder 2* and does not employ distributed source coding principles yet. A scalar quantizer is used to represent the DCT coefficients of all temporal bands. The quantized coefficients are simply run-level encoded. On the other hand, *Encoder 2* is designed for distributed source coding and uses nested lattice codes to represent the DCT coefficients of all temporal bands.

B. Nested Lattice Codes for Transform Coefficients

The 8×8 DCT coefficients of *Encoder 2* are represented by a 1-dimensional nested lattice code [6]. Further, we construct cosets in a memoryless fashion [7].

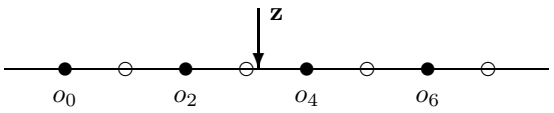

Fig. 3. Coset-coding of transform coefficients where *Encoder 2* transmits at a rate R_{TX} of 1 bit per transform coefficient.

Fig. 3 explains the coset-coding principle. Assume that *Encoder 2* transmits at a rate R_{TX} of 1 bit per transform coefficient and utilizes two cosets $\mathcal{C}_{1,0} = \{o_0, o_2, o_4, o_6\}$ and $\mathcal{C}_{1,1} = \{o_1, o_3, o_5, o_7\}$ for encoding. Now, the transform coefficient o_4 shall be encoded and the encoder sends one bit to signal coset $\mathcal{C}_{1,0}$. With the help of the side information coefficient $\mathbf{z}$, 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.

Consider the 64 transform coefficients $\mathbf{c}_i$ of the 8×8 DCT at *Encoder 2*. The correlation between the i-th transform coefficient $\mathbf{c}_i$ at *Encoder 2* and the i-th transform coefficient of the side information $\mathbf{z}_i$ depends strongly on the coefficient index i. In general, the correlation between corresponding DC coefficients ($i = 0$) is very high, whereas the correlation between corresponding high-frequency coefficients decreases rapidly. To encounter the problem of varying correlation, we adapt the transmission rate R_{TX} to each transform coefficient. For weakly correlated coefficients, a higher transmission rate has to be chosen.

Adapting the transmission rate to the actual correlation is accomplished with nested lattice codes [6]. The idea of nested lattices is, roughly, to generate diluted versions of the original coset code. As we use uniform scalar quantization, we consider the 1-dimensional lattice. Fig. 4 depicts the fine code $\mathcal{C}_0$ in the Euclidean space with minimum distance Q. $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ are nested codes with the ν-th coset $\mathcal{C}_{\mu,\nu}$ of $\mathcal{C}_\mu$ relative to $\mathcal{C}_0$. The nested codes are coarser and the union of their cosets gives the fine code $\mathcal{C}_0$, i.e. $\bigcup_\nu \mathcal{C}_{1,\nu} = \mathcal{C}_0$.

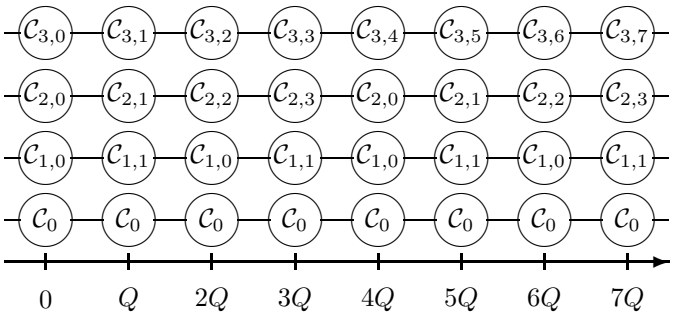

Fig. 4. Nested lattices. The 1-dimensional fine code $\mathcal{C}_0$ is embedded into the Euclidean space with minimum distance Q. $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ are nested codes with the ν-th coset $\mathcal{C}_{\mu,\nu}$ of $\mathcal{C}_\mu$ relative to $\mathcal{C}_0$.

The binary representation of the quantized transform coefficients determines its coset representation in the nested lattice. If the transmission rate for a coefficient is $R_{TX} = \mu$, then the μ least significant bits of the binary representation determine the ν-th coset $\mathcal{C}_{\mu,\nu}$. For highly correlated coefficients, the number of required cosets and, hence, the transmission rate is small. To achieve efficient entropy coding of the binary representation of all 64 transform coefficients, we define bit-planes. Each bit-plane is run-length encoded and transmitted to *Decoder 2* upon request.

C. Decoding with Side Information

At *Encoder 2*, the quantized transform coefficients are represented with 10 bit-planes, where 9 are used for encoding the absolute value, and one is used for the sign. *Encoder 2* is able to provides the full bit-planes, independent of any side information at the *Decoder 2*. *Encoder 2* is also able to receive a bit-plane mask to weight the current bit-plane. The masked bit-plane is run-length encoded and transmitted to *Decoder 2*.

Given the side information at *Decoder 2*, masked bit-planes are requested from *Encoder 2*. For that, *Decoder 2* sets the bit-plane mask to indicate the bits that are required from *Encoder 2*. Dependent on the received bit-plane mask, *Encoder 2* transmits the weighted bit-plane utilizing run-length encoding. *Decoder 2* attempts to decode the already received bit-planes with the given side information. In case of decoding error,

Decoder 2 generates a new bit-plane mask and requests a further weighted bit-plane.

Decoder 2 has the following options for each mask bit: If a bit in the bit-plane is not needed, the mask value is 0. The mask value is 1 if the bit is required for error-free decoding. If the information at the decoder is not sufficient for this decision, the mask is set to 2 and the encoded transform coefficient that is used as side information is transmitted to *Encoder 2*. With this side information $\mathbf{z}_i$ for the i-th transform coefficient $\mathbf{c}_i$, *Encoder 2* is able to determine its best transmission rate $\mu = R_{TX}[i]$ and coset $\mathcal{C}_{\mu,\nu}$. This information is incorporated into the current bit-plane and transmitted to *Decoder 2*: Bits that are not needed for error-free decoding are marked with 0. Further, 1 indicates that the bit is needed and its value is 0, and 2 indicates that the bit is needed with value 1.

Decoder 2 aims to estimate the i-th transform coefficient $\hat{\mathbf{c}}_i$ based on the current transmission rate $\mu = R_{TX}[i]$, the partially received coset $\mathcal{C}_{\mu,\nu}$, and the side information $\mathbf{z}_i$.

$$\hat{\mathbf{c}}_i = \underset{\mathbf{c}_i \in \mathcal{C}_{\mu,\nu}}{\operatorname{argmin}} \left[\mathbf{c}_i - \mathbf{z}_i\right]^2 \quad \text{given} \quad \mu = R_{TX}[i] \tag{1}$$

With increasing number of received bit-planes, i.e. increasing transmission rate $R_{TX}[i]$, this estimate gets more accurate and stays definitely constant for rates beyond the critical transmission rate $R^*_{TX}[i]$. Therefore, a simple decoding algorithm is as follows: An additional bit is required if the estimated coefficient changes its value when the transmission rate increases by 1. An unchanged value for an estimated coefficient is just a necessary condition for having achieved the critical transmission rate. This condition is not sufficient for error-free decoding and, in this case, *Encoder 2* has to determine the critical transmission rate to resolve any ambiguity.

Note that *Decoder 2* receives the coded information in bit-plane units, starting with the plane of least significant bits. With each new bit-plane, *Decoder 2* utilizes a coarser lattice where the number of cosets as well as the minimum Euclidean distance increases exponentially.

D. Disparity-Compensated Side Information

To improve the efficiency of *Decoder 2*, the side information from *Decoder 1* is disparity compensated in the image domain. If the camera positions are unknown, the coding system estimates the disparity information from sample frames. During this calibration process, the side information for *Decoder 2* is less correlated and *Encoder 2* has to transmit at a higher bit-rate. Our system utilizes block-based estimates of the disparity values which are constant for all corresponding image pairs in the stereoscopic sequence. We estimate the disparity from the first pair of images in the sequences. The right image is subdivided horizontally into 4 segments and vertically into 6 segments. For each of the 24 blocks in the right image, we estimate half-pel accurate disparity vectors. Intensity values for half-pel positions are obtained by bilinear interpolation. Assuming that the camera positions are unaltered in time, the disparity information is used in the image domain to improve the side information in the transform domain.

III. Experimental Results

For the experiments, we select the stereoscopic MPEG-4 sequences *Funfair* and *Tunnel* in QCIF resolution. We divide 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 $QP = 2$, where $Q = 2QP$. This coded version of the left view is used for disparity compensation. The compensated frames provide the side information for *Decoder 2* to decode the right view.

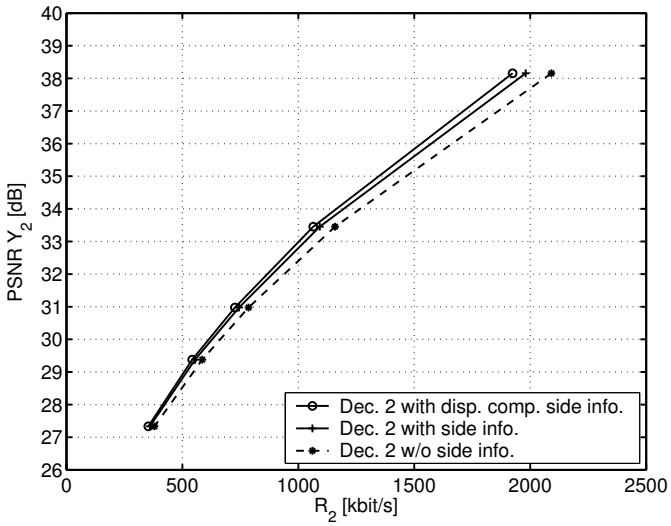

Fig. 5. Luminance PSNR vs. total bit-rate at *Decoder 2* for the sequence *Funfair 2* (right view). Compared is decoding with disparity-compensated side information, decoding with coefficient side information only, and decoding without side information. For all cases, groups of $K = 32$ pictures are used.

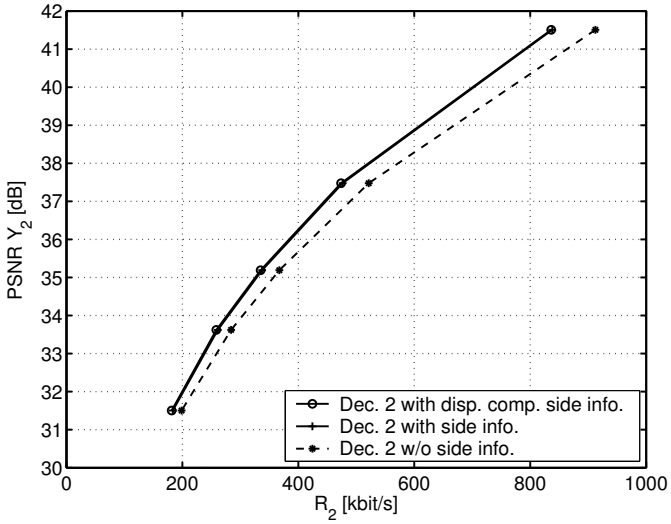

Fig. 6. Luminance PSNR vs. total bit-rate at *Decoder 2* for the sequence *Tunnel 2* (right view). Compared is decoding with disparity-compensated side information, decoding with coefficient side information only, and decoding without side information. For all cases, groups of $K = 32$ pictures are used.

Figs. 5 and 6 show the luminance PSNR over the total bit-rate of the distributed codec *Encoder 2* for the sequences *Funfair 2* and *Tunnel 2*, respectively. The sequences are the

right views of the stereoscopic sequences. The rate-distortion points are obtained by varying the quantization parameter for the nested lattice in *Encoder 2*. When compared to decoding without side information, decoding with coefficient side information reduces the bit-rate of *Funfair 2* by up to 5% and that of *Tunnel 2* by up to 8%. Decoding with disparity-compensated side information reduces the bit-rate of *Funfair 2* by up to 8%. The block-based disparity compensation has limited accuracy and is not beneficial for *Tunnel 2*. But utilizing more accurate geometrical information about the scene will improve the side information for *Decoder 2* and, hence, will further reduce the bit-rate of *Encoder 2*.

Fig. 7 and 8 show the bit-rate difference between decoding with side information and decoding without side information over the luminance PSNR at *Decoder 2* for the sequences *Funfair 2* (right view) and *Tunnel 2* (right view), respectively. The bit-rate savings due to side information are depicted for weak temporal filtering with $K = 8$ pictures per GOP and strong temporal filtering with $K = 32$ pictures per GOP. Note that both the coded signal (right view) and the side information (left view) are encoded with the same GOP length K. It is observed that strong temporal filtering results in lower bit-rate savings due to side information when compared to the bit-rate savings due to side information for weaker temporal filtering. Obviously, there is a trade-off between the level of temporal decorrelation and the efficiency of multi-view side information. This trade-off is also found in the theoretical investigation on the efficiency of video coding with side information [8].

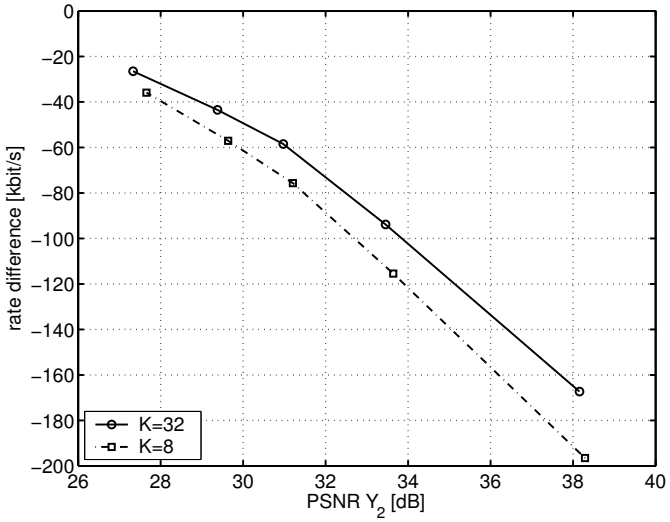

Fig. 7. Bit-rate difference vs. luminance PSNR at *Decoder 2* for the sequence *Funfair 2* (right view). The rate difference is the bit-rate for decoding with side information minus the bit-rate for decoding without side information and reflects the bit-rate savings due to decoding with side information. Smaller bit-rate savings are observed for strong temporal decorrelation ($K = 32$) when compared to the bit-rate savings for weak temporal decorrelation ($K = 8$).

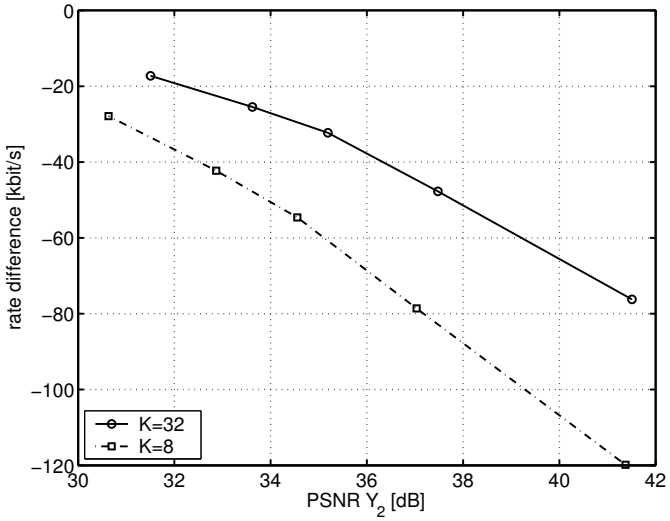

Fig. 8. Bit-rate difference vs. luminance PSNR at *Decoder 2* for the sequence *Tunnel 2* (right view). The rate difference is the bit-rate for decoding with side information minus the bit-rate for decoding without side information and reflects the bit-rate savings due to decoding with side information. Smaller bit-rate savings are observed for strong temporal decorrelation ($K = 32$) when compared to the bit-rate savings for weak temporal decorrelation ($K = 8$).

IV. Conclusions

This paper discusses distributed coding of two correlated video signals. The coding scheme is based on motion-compensated temporal wavelets and transform coding of temporal subbands. The scalar transform coefficients are represented by a nested lattice code. For this representation, we define bit-planes and encode these with run-length coding. As the correlation of the transform coefficients is not stationary, we decode with feed-back and adapt the coarseness of the code to the actual correlation. With disparity-compensated side information, we observe up to 8% bit-rate savings over decoding without side information. Finally, we observe a trade-off between the level of temporal decorrelation and the efficiency of multi-view side information.

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.

[2] S. S. Pradhan and K. Ramchandran, "Enhancing analog image transmission systems using digital side information: a new wavelet-based image coding paradigm," in *Proceedings of the Data Compression Conference*, Snowbird, UT, Mar. 2001, pp. 63–72.

[3] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, to appear, invited paper.

[4] X. Zhu, A. Aaron, and B. Girod, "Distributed compression for large camera arrays," in *Proceedings of the IEEE Workshop on Statistical Signal Processing*, St. Louis, MO, Sept. 2003.

[5] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," *Signal Processing: Image Communication*, vol. 19, 2004, to appear.

[6] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[7] 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.

[8] M. Flierl, "Distributed coding of dynamic scenes," Swiss Federal Institute of Technology, Lausanne, Switzerland, Tech. Rep. EPFL-ITS-2004.015, Jan. 2004. [Online]. Available: http:// itswww.epfl.ch/ ˜mflierl/ publications.html