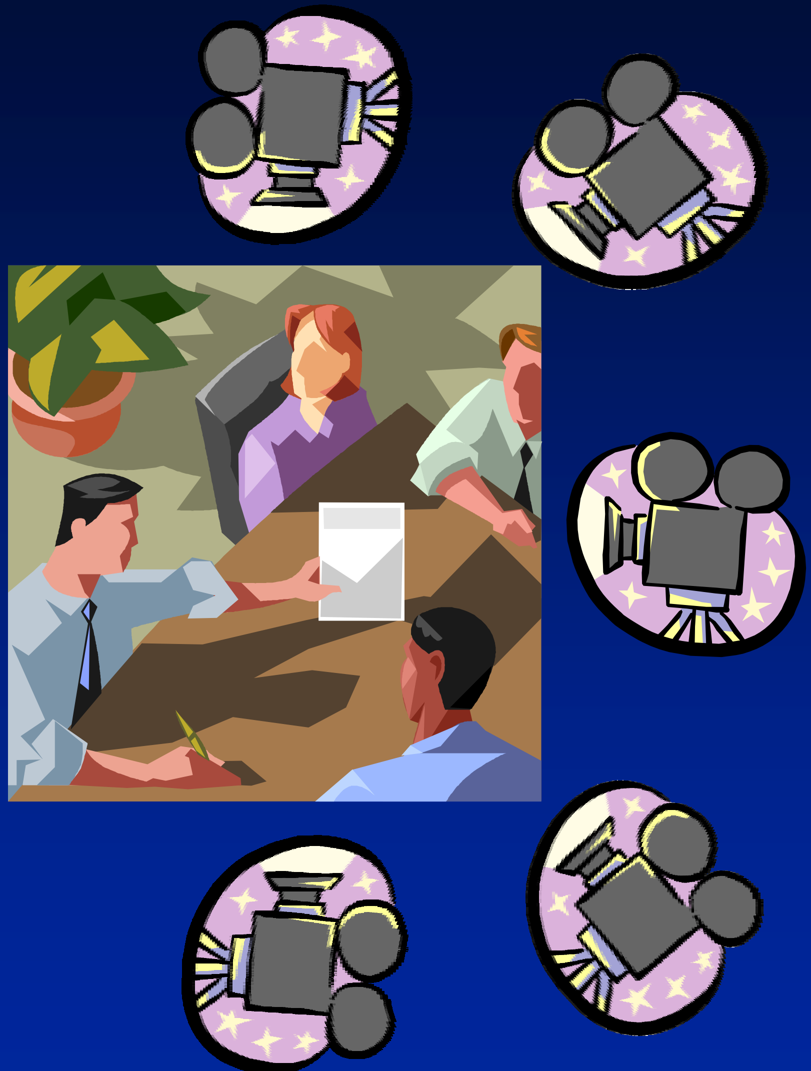
International Conference on Acoustics, Speech, and Signal Processing 2005

Video Coding with Motion-Compensated Temporal Transforms and Side Information

Markus Flierl and Pierre Vandergheynst

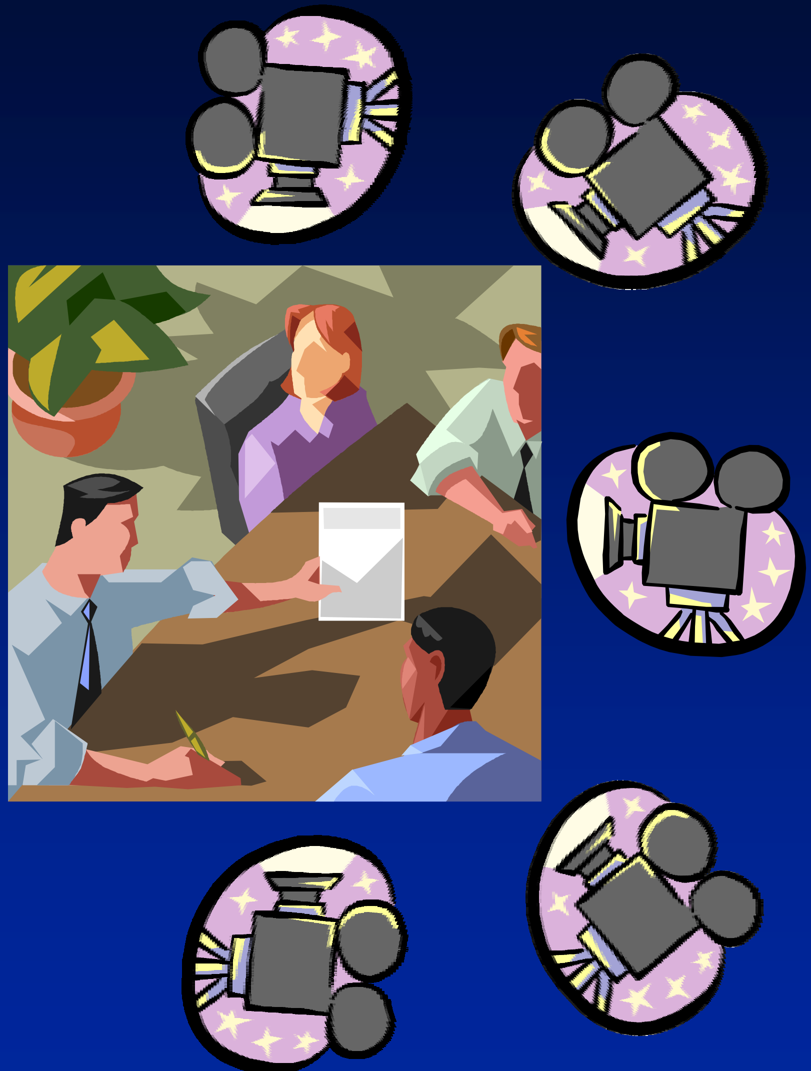
Signal Processing Institute
Swiss Federal Institute of Technology, Lausanne

Coding Scenario I



- 3-dimensional scene that evolves in time
- Observed by multiple video cameras located at different positions
- Each camera signal is coded locally
- The cameras are connected directly to the network
- One remote decoder is able to reconstruct arbitrary views

Coding Scenario II

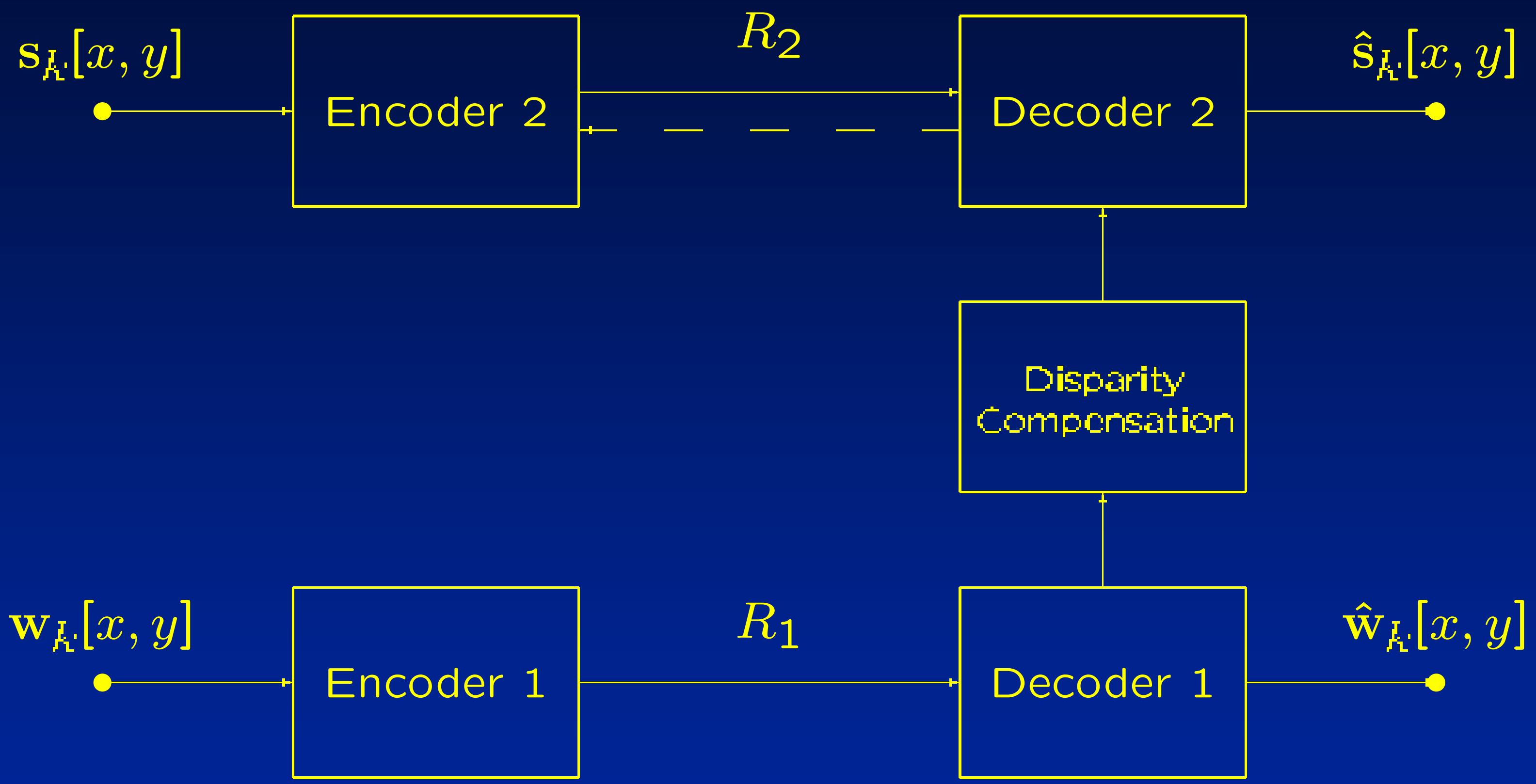


- How to use the information of neighboring cameras to improve the efficiency of the current video encoder?
- Obviously, side information can be used at encoder and decoder
- But what if the encoder do not communicate directly?
- Each encoder needs to operate independently and to transmit robustly to the central decoder!

Outline

- Coding architecture with disparity compensation at the decoder
- Rate distortion with video side information
- Signal model for subband coding of video
- Conditional Karhunen-Loeve transform
- Performance bounds
- Motion-compensated temporal Haar wavelet
- Experimental results

Coding Architecture with Disparity Compensation

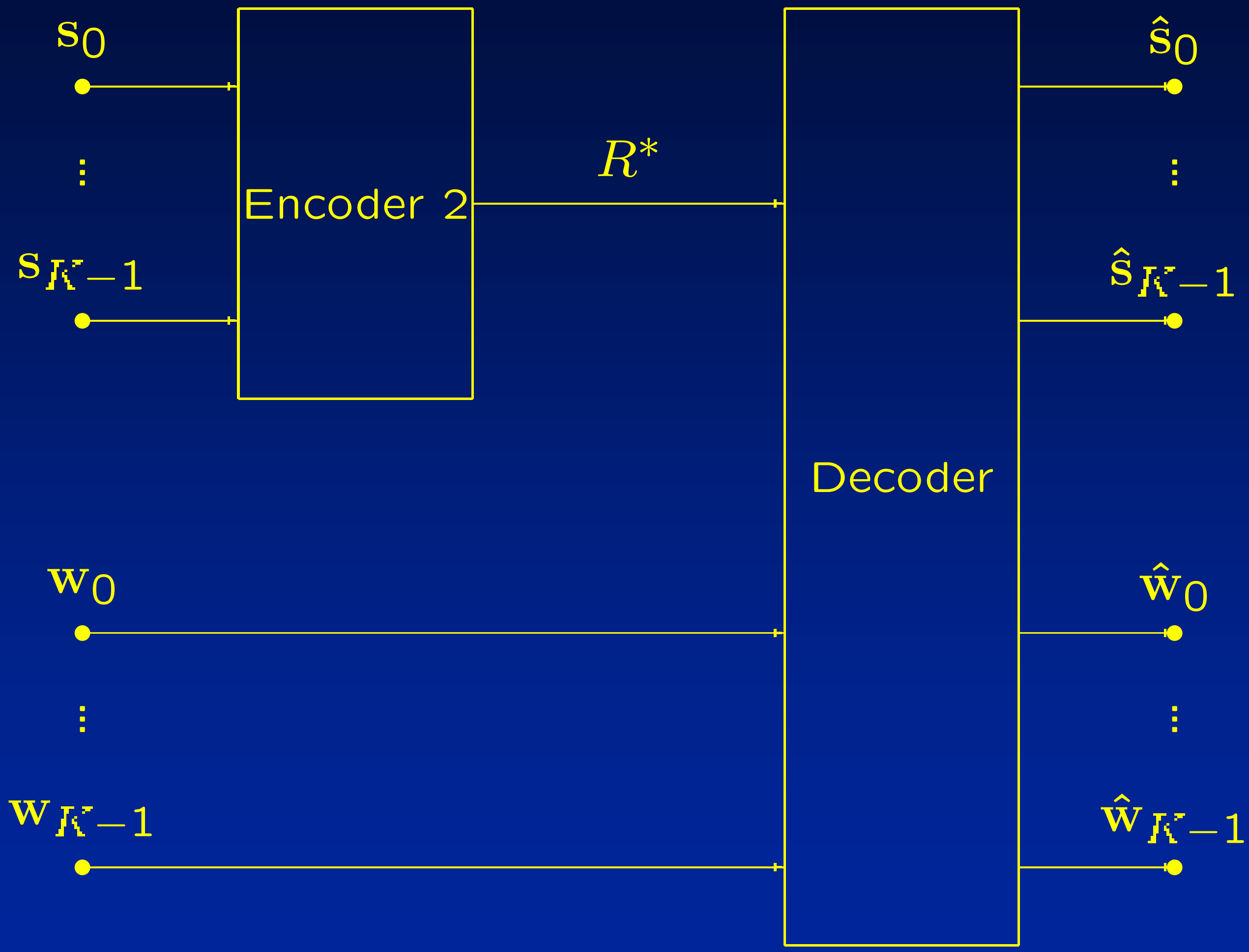


How to encode the video signal at each camera?

Rate Distortion with Video Side Information

- Vector **s** of *K* input pictures to be encoded
- Vector **w** of *K* side information pictures
- At high rates:
  - ***Reconstructed side information at the decoder approaches the original side information, i.e.,***
    $\hat{\mathbf{w}} \longrightarrow \mathbf{w}$
  - ***Wyner-Ziv coding scheme***
  - ***Rate distortion function of Encoder 2 is bounded by the conditional rate distortion function [Wyner & Ziv, 1976]***

Rate Distortion with Video Side Information



Rate Distortion with Video Side Information

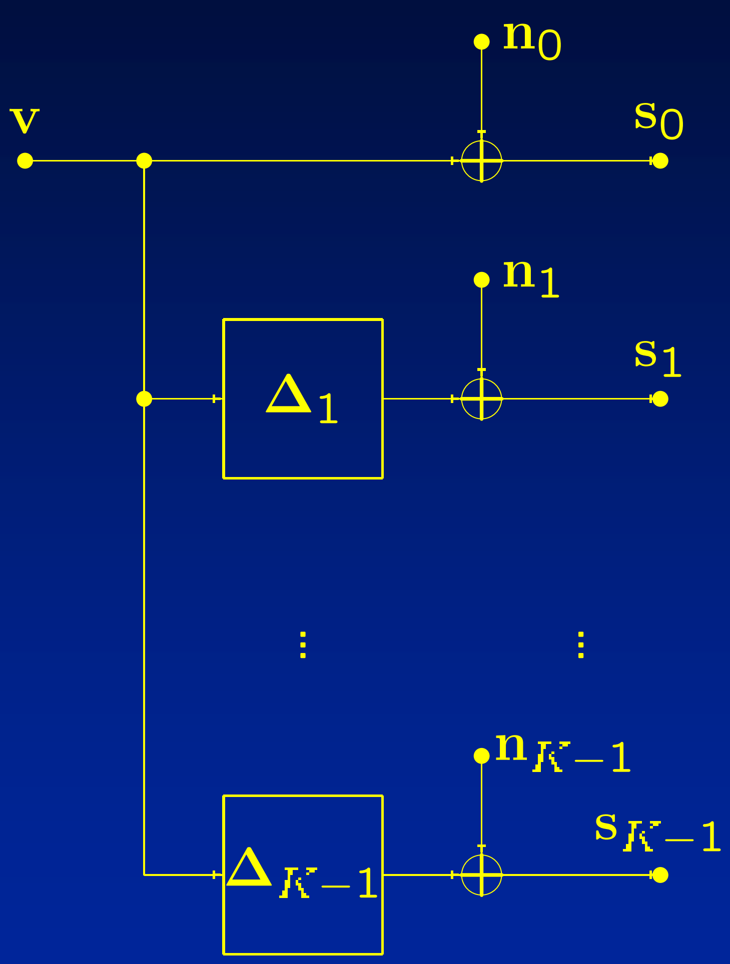
- Very accurate disparity compensation
- Consider illumination changes and occlusions
- Side information is a noisy version of the video signal to be encoded:

$$\mathbf{w} = \mathbf{s} + \mathbf{u}$$

- The vector of noisy images **u** is statistically independent of the vector of input pictures **v**.
- Matrix of conditional power spectral densities:

$$\Phi_{\mathbf{s}|\mathbf{w}}(\omega) = \Phi_{\mathbf{ss}} \left[\Phi_{\mathbf{ss}} + \Phi_{\mathbf{uu}}\right]^{-1} \Phi_{\mathbf{uu}}$$

Signal Model for Subband Coding of Video



Model for coding with motion-compensated lifted wavelets [Flierl & Girod, 2003]

$\mathbf{v}$	model picture
$\mathbf{\Delta}_k$	k-th displacement error
$\mathbf{n}_k$	k-th noise signal
$\mathbf{s}_k$	k-th motion-compensated signal

Signal Model for Subband Coding of Video

- Basic idea:
  - *Reversible true motion trajectories*
  - *Reversible estimated motion trajectories*
  - *Identical accuracy of motion compensation*
- Power spectral densities of *K* pictures:

$$\frac{\Phi_{\mathbf{ss}}(\omega)}{\Phi_{\mathbf{vv}}(\omega)} = \begin{pmatrix} 1+\alpha & P & \cdots & P \\ P & 1+\alpha & \cdots & P \\ \vdots & \vdots & \ddots & \vdots \\ P & P & \cdots & 1+\alpha \end{pmatrix}$$

$\alpha(\omega, \sigma_{\mathbf{n}}^2)$
normalized PSD
of noise

$P(\omega, \sigma_{\mathbf{\Delta}}^2)$
characteristic function
of displacement error

Conditional Karhunen-Loeve Transform

- Conditional KLT of *K* motion-compensated pictures given *K* side information pictures:
  - *First eigenvector 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*
- Independent of side information, i.e., side information is not required at the encoder
- Motion-compensated Haar wavelet meets these requirements

Performance Bounds via Conditional Eigendensities

- Rate difference with respect to coding without video side information for each picture k

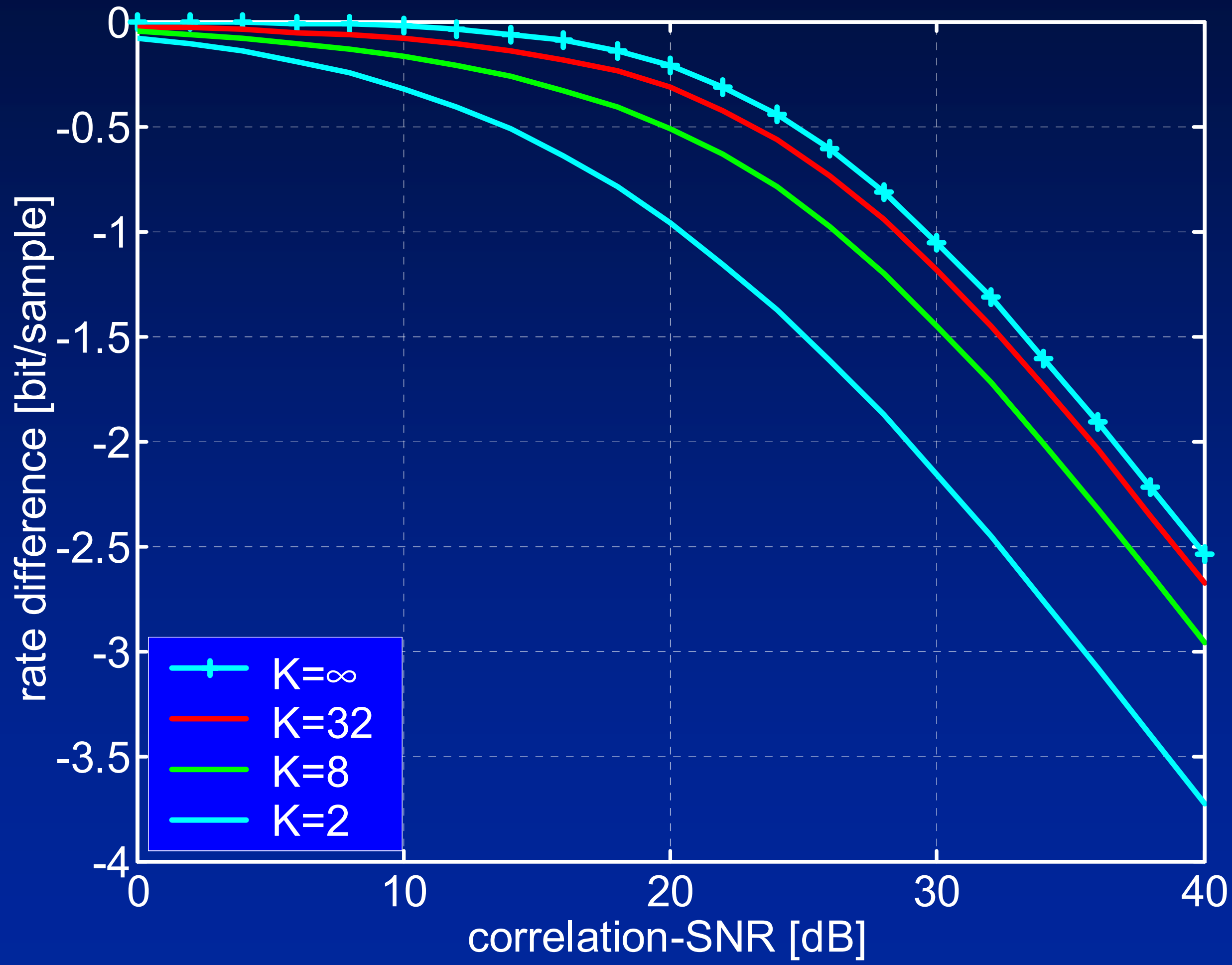
$$\Delta R_k^* = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left(\frac{\Lambda_k^*(\omega)}{\Lambda_k(\omega)} \right) d\omega$$

  - *Measures maximum bit-rate reduction*
  - *Compares to optimum coding without video side information*
  - *For the same mean squared reconstruction error*
  - *For Gaussian signals*

- Average rate difference for Encoder 2:

$$\Delta R^* = \frac{1}{K} \sum_{k=0}^{K-1} \Delta R_k^*$$

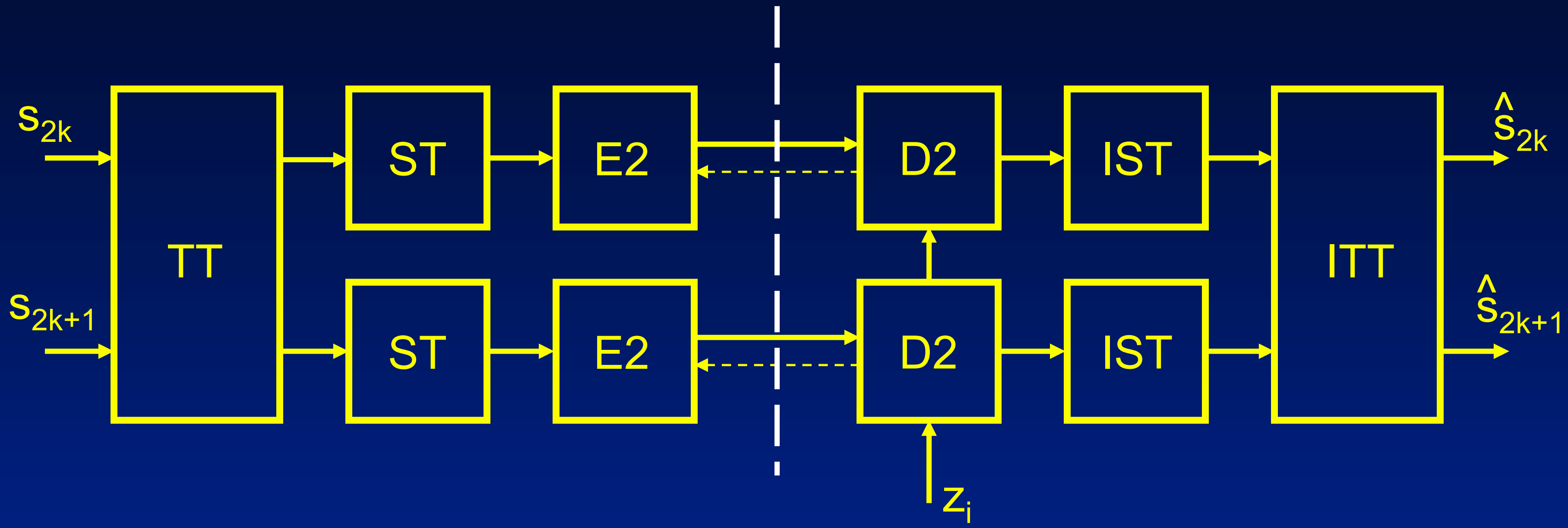
Rate Difference for High-Rate Approximation



Side information is less efficient for larger GOP sizes

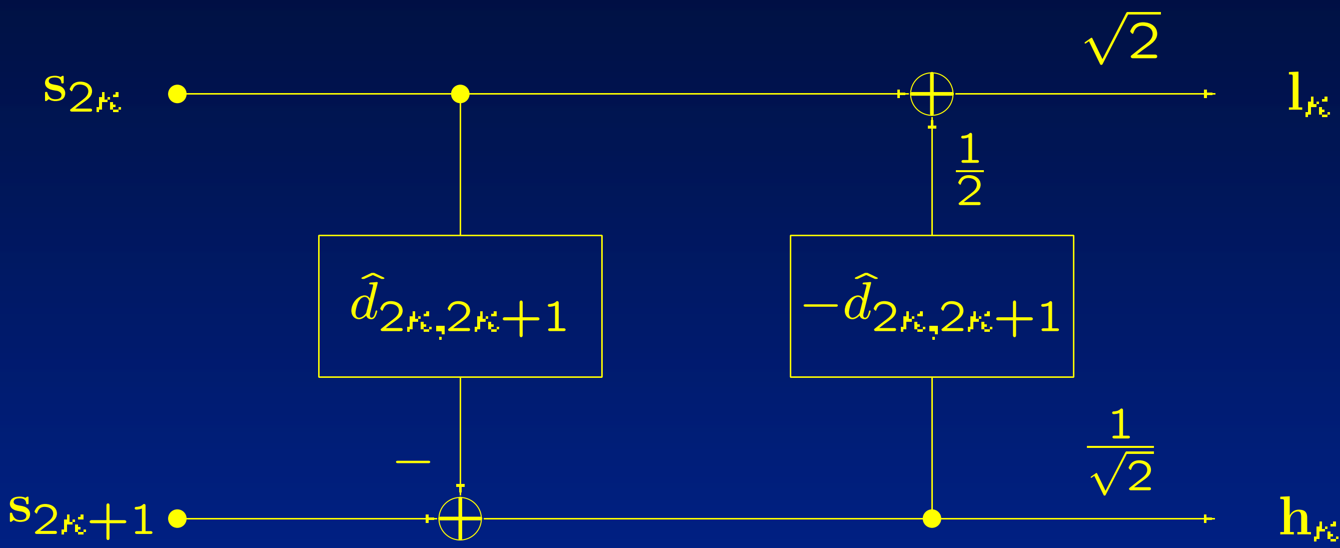
RNL = -30 dB
Half-pel accuracy

Coding Scheme



- Temporal Transform: Motion-compensated Haar wavelet
  - *Dyadic decomposition for each group of K pictures*
- Spatial Transform: 8x8 DCT
- Coefficient Coder: Nested lattice code
  - *Same minimum distance for all codes*
  - *Coefficient decoder uses side information z_i*

Motion-Compensated Temporal Haar Wavelet

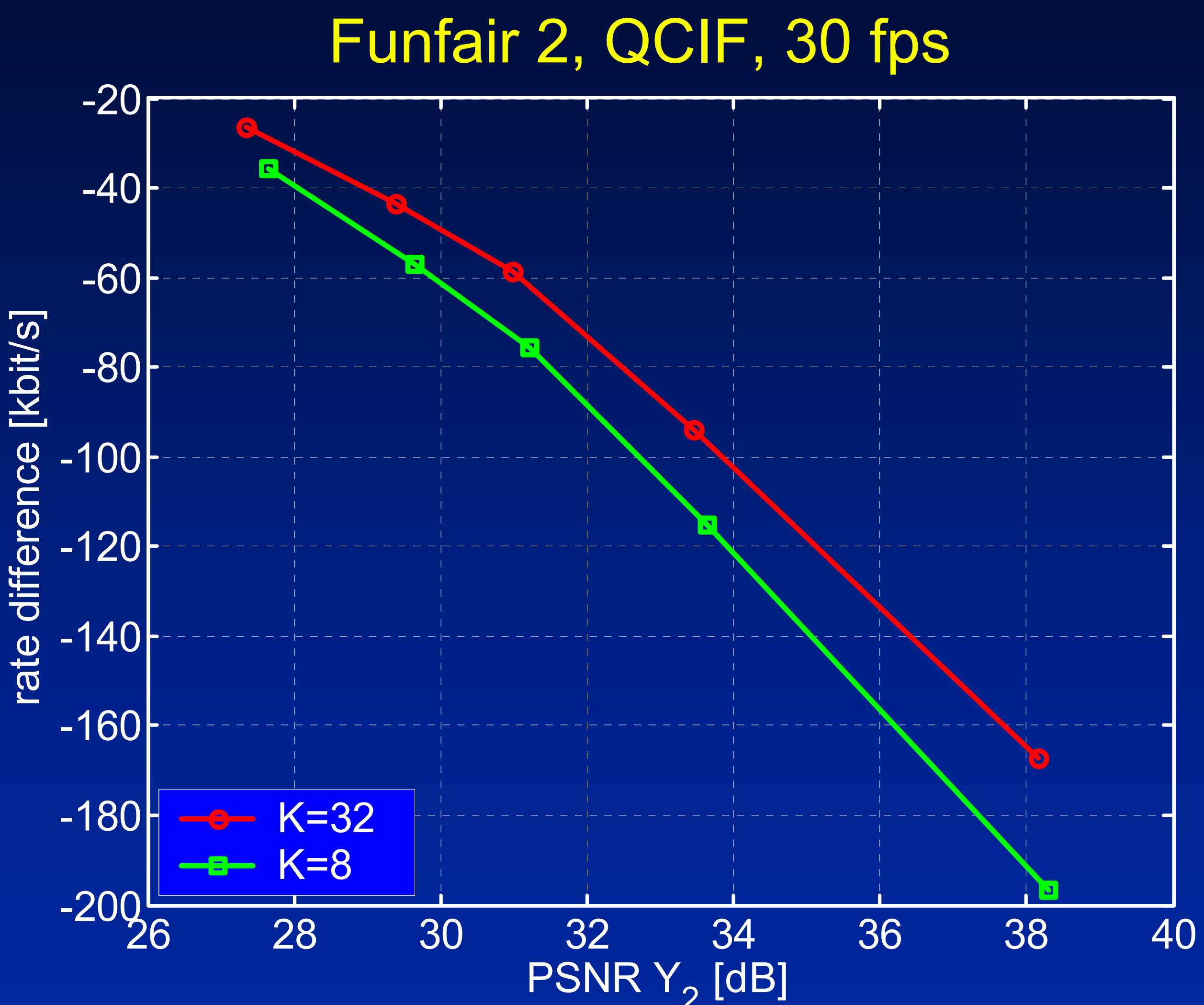


- Haar wavelet with motion-compensated lifting steps [Pesquet-Popescu & Bottreau, 2001]
- 16x16 block motion compensation with half-pel accuracy

Experimental Results

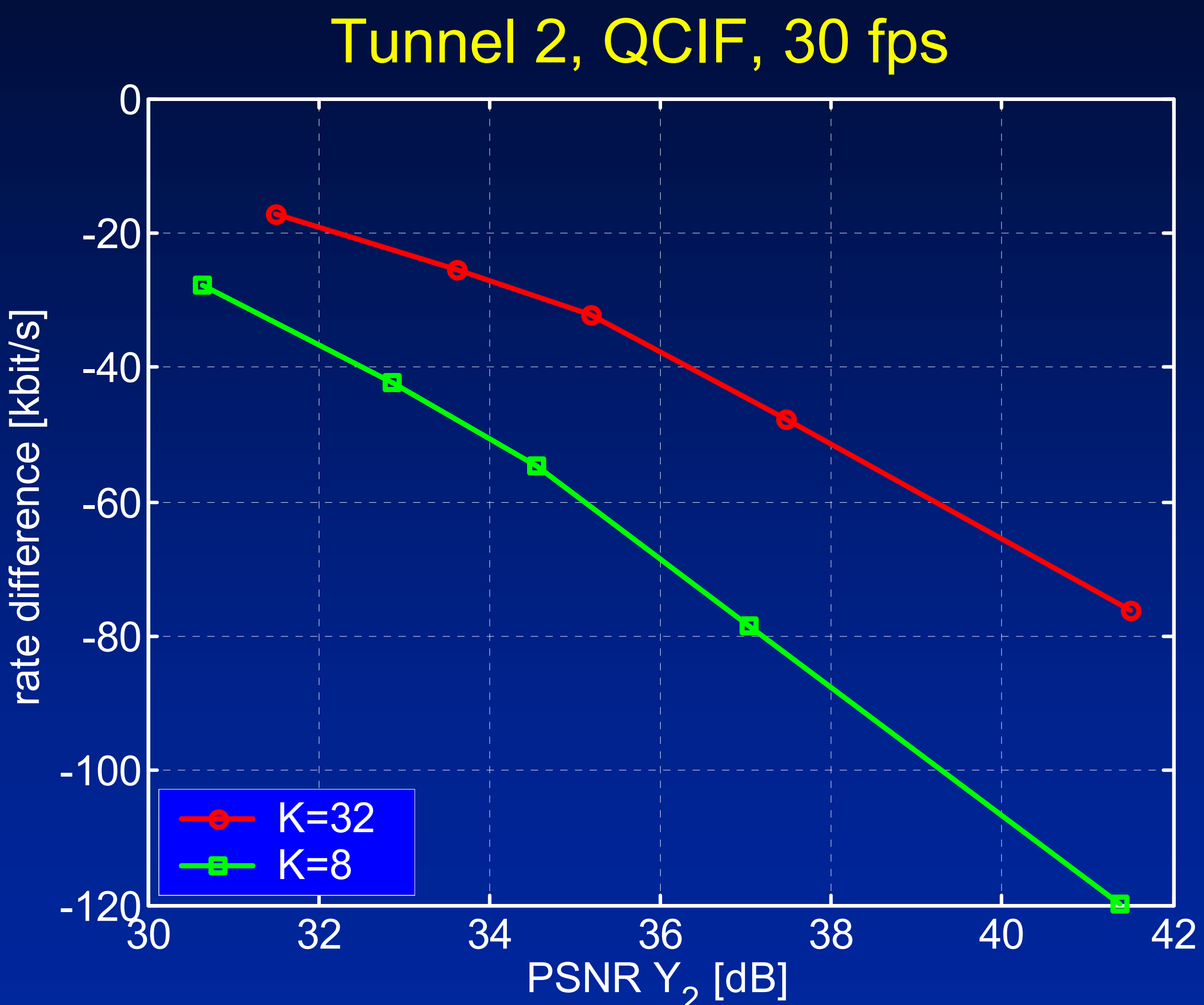
- *Encoder 1* encodes the side information (left view of a stereoscopic sequence) at high quality
- *Encoder 2* encodes the right view of a stereoscopic sequence
- The GOP sizes for *Encoder 1 & 2* are identical
- The side information is disparity compensated in the image domain
- The disparity is estimated for 24 blocks on the first frame pair
- The camera positions are unaltered in time and the disparity estimates are used for all images

Temporal Decorrelation and Video Side Information



Side information is less efficient for larger GOP sizes

Temporal Decorrelation and Video Side Information



Side information is less efficient for larger GOP sizes

Conclusions

- Robust coding of video signals in the presence of highly correlated video side information
- Rate distortion with video side information
- Conditional Karhunen-Loeve transform
  - *Motion-compensated lifted Haar wavelet*
  - *Provides a robust representation for each camera signal*
- Performance bounds via conditional eigendensities
- Observe a trade-off between the level of temporal decorrelation and the efficiency of multi-view side information:
  - *Efficiency of side information increases for decreasing GOP size*