Coding Efficiency of Video Sensor Networks

Markus Flierl

Stanford University
Max Planck Center for Visual Computing and Communication
Motivation

- 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
• Communication Scenario

• Coding of One Video Signal with Side Information

• Efficiency Study
  – Model for Transform-Coded Video Signals
  – Conditional Karhunen-Loeve Transform
  – Relative Conditional Eigendensities

• Collaborative Coding of Multiple Video Signals

• Performance with Gaussian Assumption
Communication Scenario

- Multiple video cameras that are connected to a network and encode highly correlated signals
- Joint decoder recovers arbitrary views
- Scenarios:
  - All encoders communicate with each other and compress the signals jointly
  - Encoders do not communicate with each other but rely solely on joint decoding
  - Combination of above scenarios
- At high rates, all scenarios may achieve the same rate distortion bound
Communication Scenario

N view-points

Encoder 1

$R_1$

$S_1$

Encoder 2

$R_2$

$S_{K+1}$

Encoder N

$R_N$

$S_{(N-1)K+1}$

Decoder

$\hat{S}_1$

$\hat{S}_{K+1}$

$\hat{S}_{(N-1)K+1}$

$\hat{S}_N$
We are interested in...

the efficiency of
- collaborative motion-compensated transform encoding/decoding

when compared to
- non-collaborative motion-compensated transform coding

depending on
- the number of cameras
- the size of the MCT GOP at each sensor
- the correlation among the view-point signals
Choose set $s$ of $K$ input pictures to be encoded

Set $w$ of $(N-1)K$ side information pictures

At high rates:

- Reconstructed side information at the decoder approaches the original side information, i.e.,
  $$\hat{w} \rightarrow w$$

- Wyner-Ziv coding scheme

- Rate distortion function of chosen encoder is bounded by the conditional rate distortion function and the bound is achieved for Gaussian Signals
  \[Wyner & Ziv, 1976\]
Coding One Video Signal with Side Information

\( u_1 \quad \vdots \quad u_K \)

Encoder \( \hat{u}_1 \quad \vdots \quad \hat{u}_K \)

Decoder

\( w_1 \quad \vdots \quad w_{(N-1)K} \)

\( \hat{w}_1 \quad \vdots \quad \hat{w}_{(N-1)K} \)

\((N-1)K\) pictures

\( (N-1)K \) pictures

\( R_i^* \)
Model for coding with motion-compensated lifted wavelets [Flierl & Girod, 2003]

- $v$: model picture
- $\Delta_k$: $k$-th displacement error
- $n_k$: $k$-th model error of motion compensation
- $u_k$: $k$-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:**

\[
\frac{\Phi_{uu}(\omega)}{\Phi_{vv}(\omega)} = \begin{pmatrix}
1 + \alpha & P & \ldots & P \\
P & 1 + \alpha & \ldots & P \\
\vdots & \vdots & \ddots & \vdots \\
P & P & \ldots & 1 + \alpha
\end{pmatrix}
\]

\[
\alpha(\omega, \sigma^2_n) \quad \text{normalized PSD of motion model error}
\]

\[
P(\omega, \sigma^2_\Delta) \quad \text{characteristic function of displacement error}
\]
Coding One Video Signal with Side Information

- Very accurate disparity compensation
- Consider model error $z$ of disparity compensation
- Side information is a noisy version of the video signal to be encoded:
  \[ w_\mu = u + z_\mu \]
- The set of model error images $z$ is statistically independent of the set of input pictures $u$.
- Matrix of conditional power spectral densities:
  \[ \Phi_{u|w} = \Phi_{uu} \left[ (N - 1)\Phi_{uu} + \Phi_{zz} \right]^{-1} \Phi_{zz} \]
Conditional Karhunen-Loeve Transform

- Conditional KLT of $\Phi_{u|w}(\omega)$ for $K$ motion-compensated pictures $u$ given $(N-1)K$ side information pictures $w$:
  - 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
• Compare the conditional eigendensities for collaborative coding $\Lambda_k^*(\omega)$ to the corresponding for non-collaborative coding $\Lambda_k(\omega)$.

\[
\frac{\Lambda_1^*(\omega)}{\Lambda_1(\omega)} = \frac{\gamma}{[N-1]Q + \gamma} \cdot \frac{Q + \frac{\gamma{KP}}{[N-1][Q+KP]+\gamma}}{Q + KP}, \\
\frac{\Lambda_k^*(\omega)}{\Lambda_k(\omega)} = \frac{\gamma}{[N-1]Q + \gamma}, \quad k = 2, 3, \ldots, K
\]

\[
\gamma(\omega, \sigma_Z^2)
\] normalized PSD of disparity model error

\[
Q(\omega, \sigma_n^2, \sigma_{\Delta}^2)
\] normalized PSD

\[
Q = 1 + \alpha - P
\]
Collaborative Coding of Multiple Video Signals

- Power of each camera signal is the same
- Signal of the current sensor always serves as a reference view-point for disparity compensation
- The model error of disparity compensation has the same variance independent of the current sensor/reference view-point

→ Each sensor shows the same rate distortion performance
Rate difference to non-collaborative MCT coding for each picture $k$ of the $\nu$-th sensor:

$$
\Delta R^*_{\nu,k} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \left( \frac{\Lambda^*_k(\omega)}{\Lambda_k(\omega)} \right) d\omega
$$

- Measures maximum bit-rate reduction
- Compares to optimum non-collaborative motion-compensated transform coding
- For the same mean squared reconstruction error
- For Gaussian signals

Average rate difference for each camera:

$$
\Delta R^* = \frac{1}{NK} \sum_{\nu=1}^{N} \sum_{k=1}^{K} \Delta R^*_{\nu,k}
$$
Rate Difference with Gaussian Assumption at High Rates

- $\text{K} = 32$
- $\text{RNL} = -30 \, \text{dB}$
- half-pel accuracy

![Graph showing rate difference with correlation-SNR [dB].]

- $N=2$
- $N=8$
- $N=32$
- $N=128$
Conclusions

- For $N$ video sensors, we have compared the efficiency of collaborative motion-compensated coding to non-collaborative motion-compensated coding in terms of rate difference at high rate.

- For a large number of cameras,
  - doubling the number decreases the rate difference at most by 0.5 bit per sample per camera.
  - quadrupling the number compensates the correlation-SNR at least by 6 dB.