Coding of Multi-View Image Sequences with Video Sensors

Markus Flierl and Bernd Girod

Max Planck Center for Visual Computing and Communication
Communicating Dynamic 3D Scenes

Sampled Dynamic 3D Scene

Data Network

E : Encoder

Fusion Center

at remote location
Outline

- Video sensor and fusion center
- Correlated multi-view image sequences
- Distributed coding with video sensors
  - Encoder of a video sensor
  - Decoding with side information
  - Disparity-compensated side information
- Experimental results
- Model study for stereoscopic images
Coding of Multi-View Image Sequences with Video Sensors
Fusion Center at Remote Location

...10011

network interface  decoder  renderer

Fusion Center
Each camera exploits the temporal correlation among $K$ successive pictures.
Distributed Coding with Video Sensors

Decoder of sensor signal using side information

Reconstructed reference view

Coding of Multi-View Image Sequences with Video Sensors
Encoder of a Video Sensor

8x8 DCT

Temporal Transform

Spatial Transform

Wyner-Ziv Encoder

\[ \text{dyadic decomposition of } K \text{ pictures with motion-compensated Haar wavelet} \]

\[ \text{coefficient coder uses a nested lattice code} \]
Decoding with Side Information

- **Encoder n** uses a nested lattice code and transmits $R_{TX}$ syndrome bits for each transform coefficient $c$.

- **Decoder n** decodes $R_{TX}$ syndrome bits for each transform coefficient $c$ *with feedback*:
  - **Encoder n** sends the initial $R_{TX}$ syndrome bits.
  - **Decoder n** attempts to decode the transform coefficient $c$ given the received $R_{TX}$ syndrome bits and the coefficient side information $z$.

  $$\hat{c} = \arg\min_{c \in C_{\mu, \nu}} [c - z]^2 \quad \text{given} \quad \mu = R_{TX}$$

  The $\nu$-th coset of the $\mu$-th nested lattice.

  - In case of decoding error, **Decoder n** requests further syndrome bits.
  - No decoding error beyond the critical syndrome rate.
Side information from **Decoder 1** is disparity-compensated in the image domain.
Example Test Sequence *Jungle*

[3DTV Network of Excellence]
Experimental Results

- **Jungle**
  - 256x192
  - 30 fps
  - 240 frames
  - N=8 views
  - GOP size K=8
Experimental Results

- **Uli**
  - 256x192
  - 30 fps
  - 240 frames
  - N=8 views
  - GOP size K=8

![Graph showing average PSNR Y (dB) vs. R [Mbit/s/camera]](graph)

- with disp. comp. side info. (curr.)
- with disp. comp. side info. (prev.)
- with side info.
- w/o side info.
Model Study for Stereoscopic Images

- Let the image \( u[l_x,l_y] \) be a scalar Gaussian random field
- Let \( u'(x,y) \) be its space-continuous ideal reconstruction
- Let \( w[l] \) be a shifted and noisy version of the image \( u[l] \) with the deterministic 2D real-valued shift \( \Theta_c \)
  \[
  w[l] = u'(l - \Theta_c) + n[l]
  \]
- Let \( s[l] \) be a shifted and noisy version of the image \( u[l] \) with the uncertain shift \( \Theta \), distributed with the PDF \( f_{\Theta}(\Theta) \)
  \[
  s[l] = u'(l - \Theta) + n[l]
  \]
- Compare conditional differential entropy rate differences:
  \[
  H(u|s)-H(u) \quad \text{vs.} \quad H(u|w)-H(u)
  \]

estimated white noise
Model Study for Stereoscopic Images

Image $u$: scalar Gaussian random field

Image $w$: $f_{\Theta}(\Theta) = \delta(\Theta - \Theta_c)$

Image $s$: $f_{\Theta}(\Theta) = \frac{1}{2a} 1_{[-a,a]}(\Theta)$

Note: deterministic disparity does not affect $H(u|w)$
Conclusions

- Exploit view-correlation of multi-view image sequences
- Operate video sensors in a collaborative fashion
- Centralized decoder performs disparity compensation
- Our experiments show that:
  - Disparity-compensated side information reduces bit-rate up to 10%
  - Without disparity compensation, gain is limited to 3%
- The uncertainty of the estimated disparity at the decoder causes a **entropy rate loss** when compared to centralized encoding