Hidden Screen-Camera Communication on an Invisible and Inaudible Dual Channel

Kun Qian*, Yumeng Lu*, Zheng Yang*, Kai Zhang*
Kehong Huang*, Xinjun Cai*, Chenshu Wu†, Yunhao Liu*‡

*Tsinghua University,
†University of Maryland College Park,
‡Michigan State University
Simultaneous Viewing and Communication

• Billions of videos are viewed every day and everywhere.

• Extra information are desired to be provided with videos.

TV Commercial  Video Sharing  AR Video
Hidden Screen-Camera Channel

• Extra information is embedded in the video unobtrusively.

Flicker-fusion property

High capturing rate
Our Previous Effort: ChromaCode

- Key goals of hidden screen-camera communication
  - **Invisibility**: imperceptible to human eyes.
  - **Throughput**: high data rate.
  - **Reliability**: low BER.

- Limitations
  - A throughput of 551 Kbps
  - A BER of 8%
  - Hardly meeting the requirement of the applications.
Problems of Existing Screen-Camera Communication Methods

• Problem #1: failure of screen detection
  • Rule-based algorithm
  • Learning-based algorithm

• Result
  • BER explodes with a detection error of several pixels.
Problems of Existing Screen-Camera Communication Methods

• Problem #2: failure of metadata decoding
  • Metadata is encoded in noisy video frames.

• Result
  • 24.6% frames have BERs over 8%,
  • where 29.3% are due to metadata failure.
Our Solution: AirCode

- Screen Detection
  - Tracking screen via visual odometry.
- Metadata Decoding
  - Encoding metadata in audio signals.
Screen Detection

Main Thread

Frame

Feature Extraction

Initialization

If Init

Screen Tracking

If not Init

Screen Loc

Map

KeyFrame

Screen Updating

Video Demodulation

Optimizing Thread
Initialization

• The rule-based detection algorithm is used before initialization.
• Two successfully decoded frames are used to initialize VO.
  • **Pose**: Camera’s location and orientation when each frame is captured
  • **Map**: 3D locations of feature points and *screen corner points*
Screen Tracking

• When a new frame is captured:
  • Calculate the camera pose with paired feature points.
  • Detect the screen by projecting screen corner points on the frame.

• When the decoding of the new frame fails:
  • Generate a second candidate with the rule-based algorithm.
  • Merge the screen points of the two candidates.
Screen Updating

- Selection of new keyframes
  - $\geq 0.5$ s has passed or $\leq 90\%$ common points are seen.
  - The current frame fails to be decoded with the tracking screen but can be decoded with the merged screen.

- When a new keyframe is selected
  - Jointly optimize camera poses and map points.
  - Optimize the screen corner points by minimizing the re-projection error.
Invisible and Inaudible Dual Channel

Sender

Video Frame → Texture Calculation → Video Embedding

Data Bits → Adaptive Encoding → Video Modulation

Metadata → Two-Layer Encoding → Audio Modulation

Audio Signal → Low-pass Filtering → Audio Embedding

Receiver

Screen Detection → Video Demodulation → Video Decoding

Metadata → Audio Decoding

Preamble Detection → Preamble Insertion → Audio Demodulation
Audio Control Channel

• Challenges
  • Long reverberation period (inter-symbol interference)
  • Strong spectral leakage (audible noises)
  • Frequency selectivity of audio devices (decoding errors)
Audio Packet Design

- Solution
  - Reverberation
    - Longer packet duration
    - Separated preambles and data symbols in the frequency band
  - Spectral leakage
    - Tapered-cosine window
  - Frequency selectivity
    - Manchester coding on adjacent subcarriers.
Video Frame Design

• Metadata is transmitted via audio packets.
Evaluation

• Data transmission
  • raw throughput, data goodput and BER

• User study
  • Subjective scores for insensitivity

Representative primary video clips and tags used in the experiments
Overall Performance

- AirCode achieves a consistent throughput of 1 Mbps, and an average BER of 5%, outperforming ChromaCode and InFrame++.
- The reliable audio channel maintains a BER consistently below 1%.
Performance of Screen Tracking

- AirCode detects the screen correctly in over 97% frames across different types of videos.
- All tracking errors of AirCode are within 9 pixels.
Ablation Study

• Screen detection
  • Accurate VO-based screen tracking.

• Audio control channel
  • Metadata area is saved for conveying data
  • Fewer metadata bit errors with the audio channel.

<table>
<thead>
<tr>
<th></th>
<th>Throughput (kbps)</th>
<th>BER (%)</th>
<th>Goodput (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen Detection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tracking</td>
<td>1086.4</td>
<td>6%</td>
<td>139.5</td>
</tr>
<tr>
<td>Rule</td>
<td>1073.8</td>
<td>8.3%</td>
<td>77.3</td>
</tr>
<tr>
<td>Control Channel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio</td>
<td>1084.3</td>
<td>4.4%</td>
<td>149.1</td>
</tr>
<tr>
<td>Video</td>
<td>893.7</td>
<td>4.6%</td>
<td>144.3</td>
</tr>
</tbody>
</table>
Conclusion

• We present AirCode, the first system that achieves considerably high data rates of >1Mbps.
  • VO-based screen detection.
  • Invisible video and inaudible audio dual channel.

• We prototype AirCode and evaluate it by experiments.
  • Higher throughput
  • Lower BER
  • Comparable invisibility.
Thanks!

Q&A

qiank10@gmail.com
https://kunqian.info/