# Hardware Synchronization of Multiple Kinects and Microphones for 3D Audiovisual Spatiotemporal Data Capture

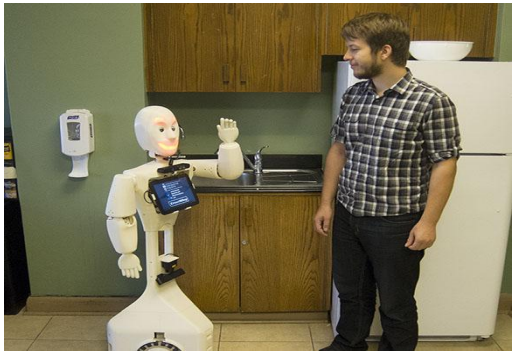
[1]Yijun Jiang, [1]David Russell, [2]Timothy Godisart, [1]Natasha Kholgade Banerjee, [1]Sean Banerjee

[1]Clarkson University, Potsdam, NY
[2]Oculus Pittsburgh, Pittsburgh, PA

# 3D Human Data Is Impacting Many Fields

Assistive robotics
- Understand how to interact with humans
- Inform movement patterns

Kinesiology
- Assist with physical therapy
- Analyze athletes, performers



http://asblab.mie.utoronto.ca/research-areas/assistive-robotics

https://www.houstonphysicianshospital.com/physical-therapy/

https://news.nike.com/news/a-look-inside-nike-s-sport-research-lab

# Markered Motion Capture Is Limiting

Solely sparse keypoints
Impedes natural actions

https://www.youtube.com/watch?v=zbkv9OxcFrg

http://www.cgarena.com/newsworld/vicon-usc-motion-capture.php

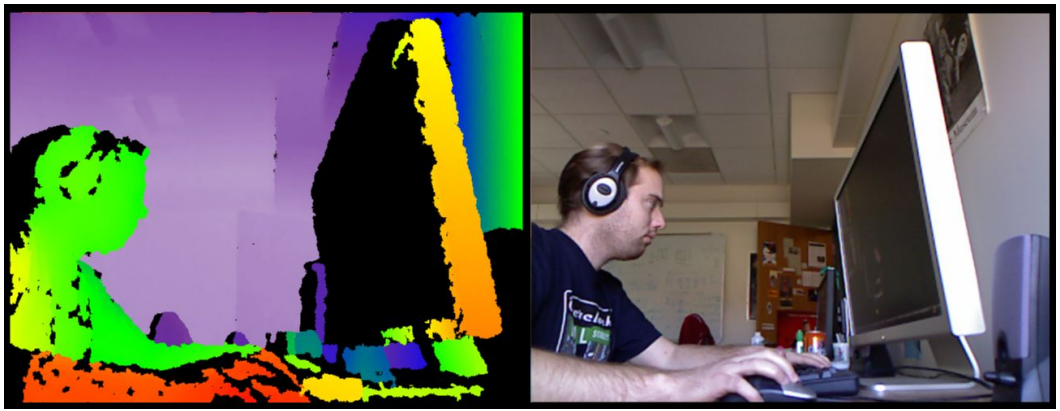
# Marker-Free Approaches Are Needed

Captures fine details
Reconstructs unstructured scenes
Allows fine motor control

# The Kinect Is A Useful Sensor

A robust color and depth camera

- Enables scene reconstruction from a few devices
- Sophisticated API



https://graphics.stanford.edu/~mdfisher/Kinect.html
/


https://en.wikipedia.org/wiki/Kinect

# However, The Kinect Has Limitations

The Kinect is a consumer-targeted device

- ➢ One capture computer per device (v2)
- ➢ No external frame trigger
- ➢ Leads to temporal offsets and ghosting artifacts

# Related Work

Network-based approaches

- Kowalski et. al., Capture requests
- Soleimani et. al., Network Time Protocol (NTP)

Environmental Cues

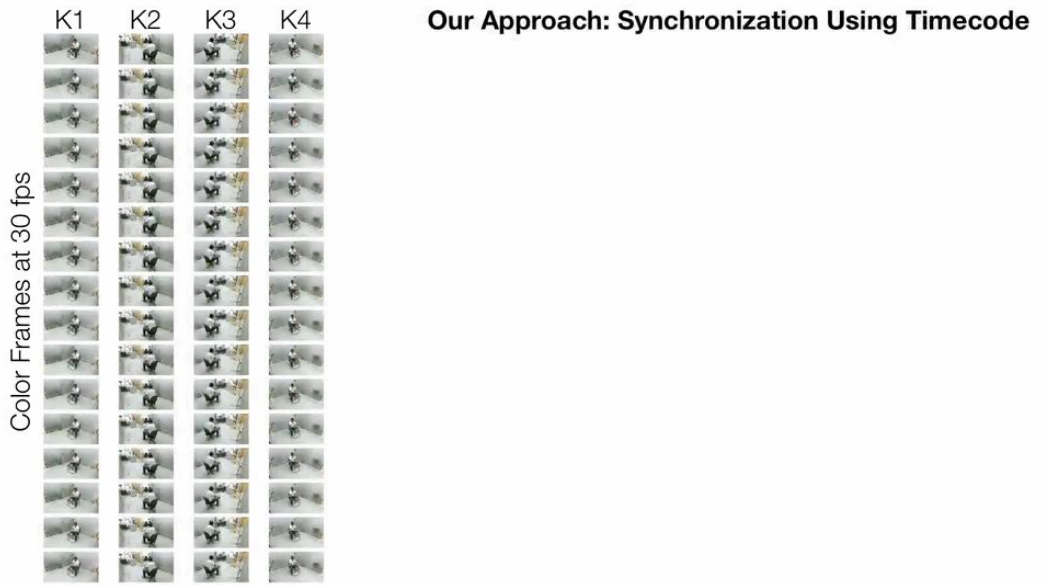
- Alexiadis et. al., Audio and visual cues

No prior approach allows synchronization of highly different modalities such as depth, RGB, audio, and thermal


- Marek Kowalski, Jacek Naruniec, and Michal Daniluk, “Live Scan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors,” in IEEE 3DV, 2015.
- Vahid Soleimani, Majid Mirmehdi, Dima Damen, Sion Hannuna, and Massimo Camplani, “3d data acquisition and registration using two opposing kinects,” in IEEE 3DV, 2016.
- Dimitrios S Alexiadis, Anargyros Chatzitofis, Nikolaos Zioulis, Olga Zoidi, Georgios Louizis, Dimitrios Zarpalas, and Petros Daras, “An integrated platform for live 3d human reconstruction and motion capturing,” IEEE TCSVT, 2017.

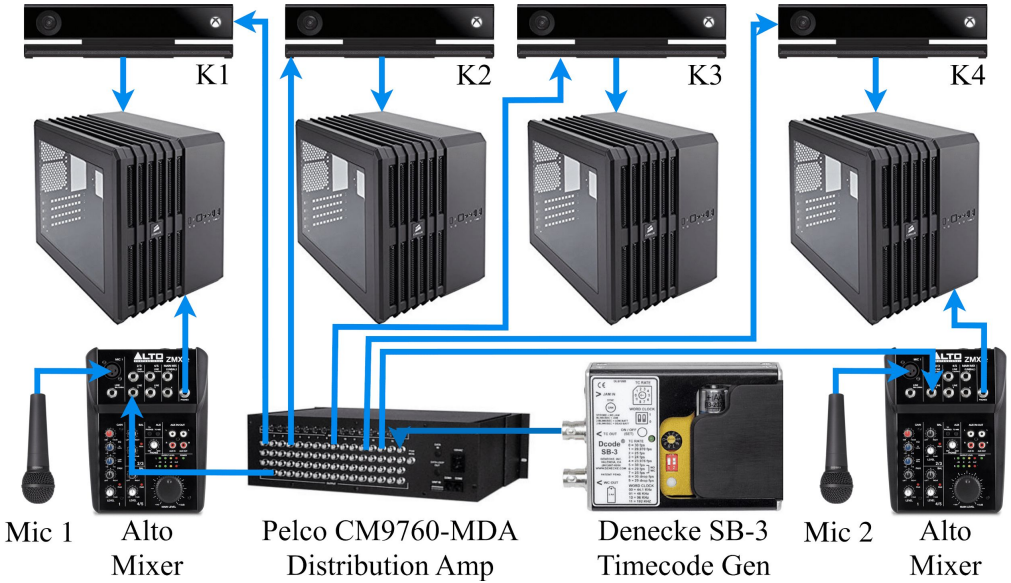
# Our Approach Utilizes Ubiquitous Audio Inputs

We feed linear timecode (LTC) into the Kinects microphone inputs
This facilitates frame-level accuracy and full-frame rate capture



# System Architecture

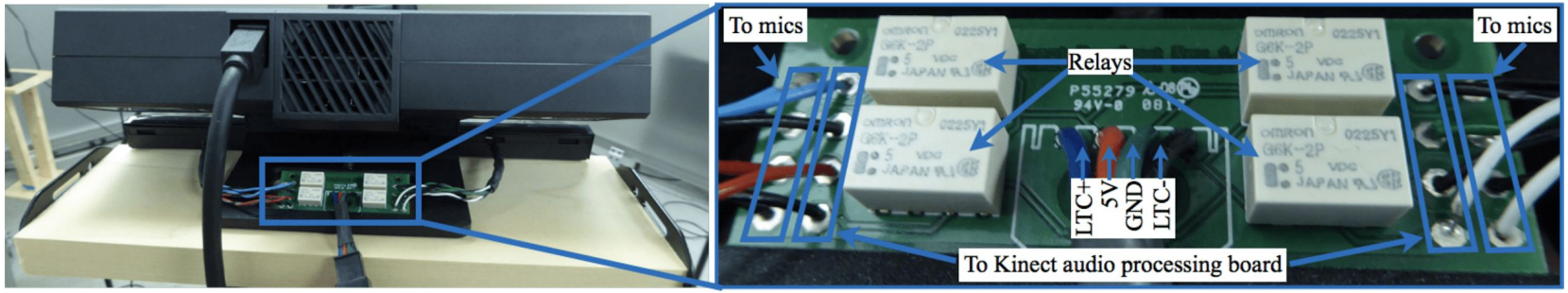
We use four Kinects and two microphone, but the approach is highly scalable



# A Compact Relay Board Injects Timecode

Our novel method utilizes the built-in microphone to receive LTC

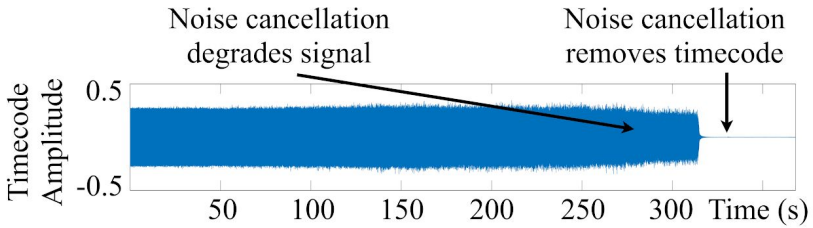
The custom circuit board injects either timecode or microphone data



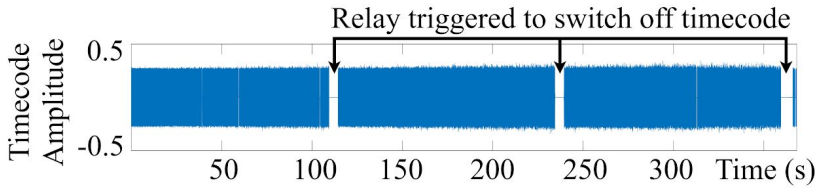
# Addressing Noise Cancellation

The timecode signal triggers active noise cancellation

- This cannot be disabled in the SDK
- We circumvent this by interspersing periods of microphone data




(a) Test Capture Showing Lack of Relay Triggering



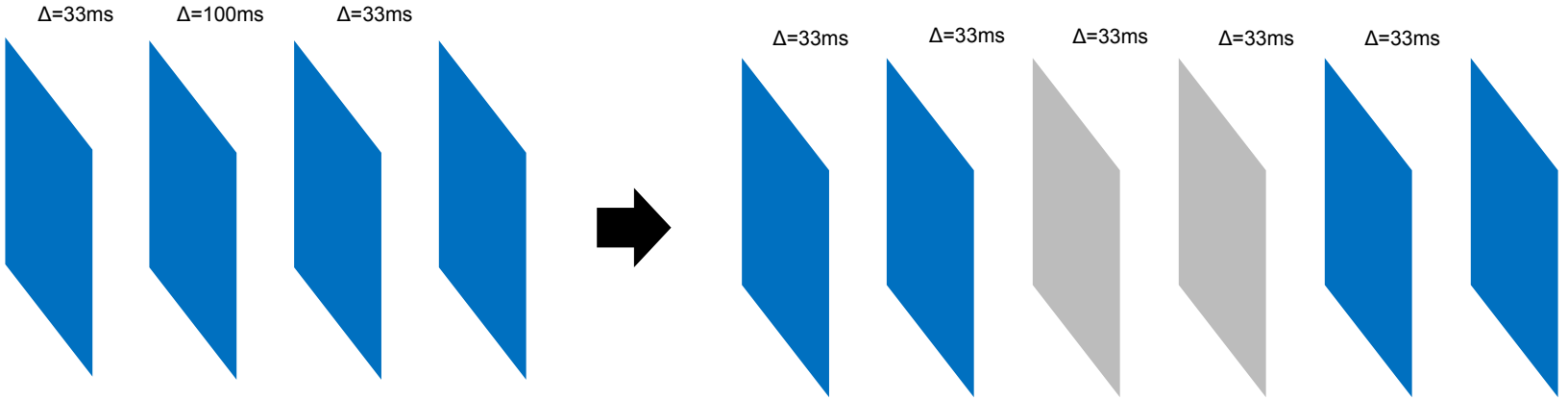

(b) Origami Capture Showing Relay Triggering

# Addressing Dropped Frames

Occasionally there will be missing frames

- For timecode, this can be caused by relay cycling
- For color, it can be operating system bottlenecks

The internal timestamp allows us to determine if the time between frames is too large



# Decoded Timestamps Should be Consecutive

Timestamp values may be improperly decoded

We enforce that each decoded value is one frame ahead of the previous one

- If not, we insert the corrected value

# Timestamping the Color Frames

Each timecode is associated with an audio frame time

We match each color frame to the nearest audio frame and assign it the audio frame’s timestamp

# Offsets in Frame-Time are Sufficiently Small

The Kinects start at different times and this causes slight misalignment
However, we show that this is consistently sub-frame

| Study Name | K1 Offset (ms) | K2 Offset (ms) | K3 Offset (ms) | K4 Offset (ms) |
|---|---|---|---|---|
| paper-plane | 14.89 ± .48 | -2.30 ± .49 | 1.02 ± .53 | 24.65 ± .42 |
| tea | 10.86 ± .47 | 26.60 ± .53 | -2.16 ± .53 | 21.09 ± .36 |
| rps | 28.53 ± .53 | 9.87 ± .42 | 6.88 ± .41 | .83 ± .48 |
| ping-pong | 26.30 ± .53 | 7.36 ± .43 | 4.63 ± .49 | -1.82 ± .41 |
| chair contest | 20.56 ± .48 | 6.81 ± .53 | -.31 ± .52 | 16.01 ± .53 |
| guitar | 22.52 ± .49 | 14.71 ± .53 | 8.13 ± .53 | 19.96 ± .51 |
| table-build | 5.32 ± .53 | 26.53 ± .53 | 10.43 ± .43 | 22.73 ± .49 |

# Integrating Microphones

We need to synchronize the high-quality microphones to the Kinects

A stereo signal contains both LTC and audio data

# Results

Here, we capture unstructured blocks

# Results

The depth cameras allow us to reconstruct thin, low-texture origami paper

# Conclusions

We present an approach to synchronize multiple Kinect v2 sensors and microphones, which avoids network latencies and environmental degradation

Future work

- Synchronize other consumer electronics with the audio jack
- Increase the quality of reconstructions with more Kinects