DIY Multiview Camera System: Panoptic Studio Teardown

Capture Software, Storage, Calibration, and Multiple Kinect Subsystem

Hanbyul Joo and Tomas Simon

Robotics Institute
Carnegie Mellon University
The Panoptic Studio
Modularized Design with 20 Panels

- 480 VGA Cameras
- 31 HD Cameras
- 10 Kinects

VGA Camera
HD Camera
VGA Camera
Projector
The Panoptic Studio

VGA (480)

HD (31)
## Synchronized Videos from 521 Views

480 VGAs, 31 HDs, and 10 RGB+Ds

### 531GB/min

<table>
<thead>
<tr>
<th>VGA (24 cameras)</th>
<th>184 MB/s</th>
<th>x 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>640 x 480 x 1 Byte/pixel x 24 cameras x 25 frame/sec</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HD</th>
<th>120 MB/s</th>
<th>x 31</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920x1080 x 2 Byte/pixel x 30 frames/sec</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How to capture?
How to store?
Capture Software
Capture Software
Why Is This Not Straightforward?

A Machine

Camera
Capture Software
Why Is This Not Straightforward?

Master Node

Clock System

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)
Capture Software

Why Is This Not Straightforward?

Local Node

Master Node

Clock System

24 VGA Cameras

480 VGA Cameras (20 Panels)

Scalability
Capture Software

Why Is This Not Straightforward?

Scalability

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)

Master Node

Clock System

Local Node

31 HD Cameras

(31 Panels)

Kinect Master Node

5 DLP Projectors

(5 Panels)
Capture Software
Why Is This Not Straightforward?

24 VGA Cameras
480 VGA Cameras (20 Panels)

31 HD Cameras
(31 Panels)

10 Kinects

Master Node
Clock System
Local Node

Diversity
Scalability

Kinect Master Node
5 DLP Projectors
(5 Panels)
Capture Software

Option 1: Asynchronized Initiation Without Preview
Capture Software
Option 1: Asynchronized Initiation Without Preview

Master Node
00:00 00:00

Clock System

Local Node

31 HD Cameras
(31 Panels)

Camera 1

Camera 2

Camera 1

Master Node

Clock System

Local Node

31 HD Cameras
(31 Panels)
Capture Software

Option 1: Asynchronized Initiation Without Preview

Thread 1

Camera 1

Camera 2

Thread 2

Time
Capture Software
Option 1: Asynchronized Initiation Without Preview

Thread 1
Thread 2

Camera 1
Camera 2

Frame 0

Time
We cannot see the data during the capture
Capture Software
Option 2: Synchronized Initiation With Preview

Master Node
Clock System

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)
Capture Software
Option 2: Synchronized Initiation With Preview

Local servers

Local Node

Clock System

Master Node

24 VGA Cameras

480 VGA Cameras (20 Panels)
Capture Software
Option 2: Synchronized Initiation With Preview

Local servers

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)

Preview streaming

Master Node

Clock System

Local servers

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)
Capture Software
Option 2: Synchronized Initiation With Preview

Local servers

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)

Master Node

Clock System

00:00

Preview streaming

Capture start-end time
• Pre-visualization
• Aligned start time
Capture Software
Take Home Messages

Timestamped frames allow a posteriori time alignment

Better capture software may provide conveniences (e.g., preview)
Data/Storage
Data / Storage
Why Do We Need To Consider This?

We would prefer,

- Higher resolution
- Higher frame rate
- More viewpoints

In Panoptic Studio,
8.85 GB/sec = 531 GB/min
Data / Storage

Options?

- Do not save images
- Real-time system

- Save compressed data

- Save raw data
Data / Storage
Our Solution

- 24 VGA Cameras
- 480 VGA Cameras (20 Panels)
- 31 HD Cameras (31 Panels)
- 10 Kinects
- 5 DLP Projectors (5 Panels)

Clock System
Master Node
Kinect Master Node

Our Solution
Data / Storage
Our Solution: Local Storage Systems

Clock System
Master Node

24 VGA Cameras
480 VGA Cameras (20 Panels)

Local Storage

31 HD Cameras
(31 Panels)

Local Storage

Kinect Master Node

10 Kinects

5 DLP Projectors
(5 Panels)

Scalability
Diversity

Local Storage Systems

Diversity
Scalability
Data / Storage

How Big Is The Panoptic Studio Data?

VGA (24 cameras): **184 MB/s × 20**
640 × 480 × 1 Byte/pixel × 24 cameras × 25 frame/sec

HD: **120 MB/s × 31**
1920x1080 × 2 Byte/pixel × 30 frames/sec

Kinect: **145 MB/s × 10**
1920 × 1080 × 2 Byte/pixel × 30 fps
+ depth 512 × 424 × 2 Byte/pixel × 30 fps (12.4 MB/s)
+ IR 512 × 424 × 2 Byte/pixel × 30 fps (12.4 MB/s)
+ body keypoints, audio

Note: average HDD’s writing speed: **80-160 MB/s**
Data / Storage
How Big Is The Panoptic Studio Data?

VGA (24 cameras): 184 MB/s x 20
640 x 480 x 1 Byte/pixel x 24 cameras x 25 frame/sec

Why not SSD?
SSD (1TB) can capture 90 minutes

HD: 120 MB/s x 31
1920x1080 x 2 Byte/pixel x 30 frames/sec

Kinect: 145 MB/s x 10
1920 x 1080 x 2 Byte/pixel x 30 fps
+ depth 512 x 424 x 2 Byte/pixel x 30 fps (12.4 MB/s)
+ IR 512 x 424 x 2 Byte/pixel x 30 fps (12.4 MB/s)
+ body keypoints, audio

Note: average HDD’s writing speed: 80-160 MB/s
Storage System
Long-Term Storage

- NAS 12 x 8TB (Synology)
- 1 NAS: 88 TB (with RAID5) = 166 minutes
- $29 / min
  $1,700 + $260 x 12 = $4,820 / 166 min
- Currently Panoptic Studio has about 1 PB data
Storage System
Take Home Messages

Data size vs Writing speed

Required capacity

Capture storage ≠ Long-term storage
Camera Calibration
Camera Calibration

Why Do We Need This?

Intrinsic: \( K \)
Extrinsic: \( R \ t \)
Lens distortion: \( k_1, k_2, p_1, p_2, p_3 \)
Camera Calibration
Space of Options

2D Checker Board  Wand (Light, Ball)  3D Pattern  Structure from Motion
Camera Calibration
A Single Camera Example

Intrinsic: $\mathbf{K}$
Extrinsic: $\mathbf{R}, \mathbf{t}$
Lens distortion: $k_1, k_2, p_1, p_2, p_3$
Camera Calibration
A Single Camera Example

Intrinsic: $K$
Extrinsic: $R$, $t$
Lens distortion: $k_1, k_2, p_1, p_2, p_3$
A Planar Pattern To Calibrate Panoptic Studio?

- **Accuracy**
  - Good accuracy

- **Scalability**
  - Need to cover many locations

- **Diversity**
  - Problem for sensors with different frame rates (VGA 25fps, HD 30fps, Kinects 30fps)
Our Solution For Camera Calibration
How To Calibrate Panoptic Studio Cameras

- Different types of cameras
  - No perfect sync among different types
  - Pattern should be stationary

- A large working volume
  - Pattern should cover as much space as possible

- 500+ cameras
  - Fully automatic method is needed
  - Avoid any image selections
A 3D Calibration Structure
using Structure from Motion (SfM)
Panoptic Studio Camera Calibration
We Use A Tent
Panoptic Studio Camera Calibration

We Use A Tent
Panoptic Studio Camera Calibration

We Use A Tent by Projecting A Random Pattern
Panoptic Studio Camera Calibration

We Use A Tent by Projecting A Random Pattern

VGA  HD  Kinect Color
Calibration Pipeline
Based on Structure from Motion (SfM)

Capture → Matching → SfM → Bundle Adjustment

Intrinsic: $K$
Extrinsic: $R, t$
Lens distortion: $k_1$

Intrinsic: $K$
Extrinsic: $R, t$
Lens distortion: $k_1, k_2, p_1, p_2, p_3$
Based on Structure from Motion (SfM)

Capture 1

Matching

Capture 2

Matching → Merging → SfM

Capture 3

Matching

Intrinsic: $K$
Extrinsic: $R \ t$
Lens distortion: $k_1$

Intrinsic: $K$
Extrinsic: $R \ t$
Lens distortion: $k_1, k_2, p_1, p_2, p_3$

Bundle Adjustment
Calibration Pipeline
Based on Structure from Motion (SfM)

Capture 1 ➔ Matching ➔ Merging ➔ SfM ➔ Bundle Adjustment

Capture 2 ➔ Matching

Capture 3 ➔ Matching

Intrinsic: $K$
Extrinsic: $R \cdot t$
Lens distortion: $k_1, k_2, p_1, p_2, p_3$
Panoptic Studio Camera Calibration

Discussion

- Less accurate keypoint localization
- Scale factor is missing
- 5 minutes for capture
- Fully automatic process
- Applicable for different types of sensors (except Kinect depth)
Camera Calibration
Take Home Messages

Three Basic Steps for Calibration:

- Identify & match correspondences
- Initialization (camera matrices and 3D points)
- Bundle adjustment

• Keypoint localization
• Coverage of working volume
• Practicality

• VisualSfM + custom codes
• COLMAP (https://colmap.github.io/)
Miscellaneous
Audio Capture
Timestamped by HD Cameras

Master Node

Clock System

31 HD Cameras (31 Panels)

Currently 15 wireless + 3 fixed microphones
Audio Capture
Timestamped by HD Cameras

Master Node

Clock System

31 HD Cameras (31 Panels)

LTC

Audio

stereo audio
Lighting

We originally didn't plan for this
• Want to get brightness with a cheap solution
• Floor lights to reduce shadow issue
• Important in high frame rate cameras
Multiview RGB-D: Kinect Subsystem
The Panoptic Studio

- Kinect v2 (10)
- VGA (480)
- HD (31)
The Kinoptic Studio

10 Kinect Subsystem

Kinect v2 (10)

Depth/Infrared

Color
RGB-D Sensor Input
RGB-D Sensor Input
Ten RGB-D Sensors
Sensor Placement
Sensor Placement
Spatial Sampling

~50K points

One ~2M points Five ~10M points Ten ~20M points
Synchronizing Diverse Sensors
Set Signals on a Common Timebase

Time (ms)

VGA
HD
Kinect

1/25
1/29.97
1/30

67
Synchronizing Diverse Sensors
Set Signals on a Common Timebase

Time (ms)
Spatio-Temporal Sampling
Fast Motion Artifacts
SMPTE standard, many decoders exist:

http://x42.github.io/libltc/

Decoding LTC Timecode

From production cameras 1/29.97

Starting sample

04:10:54:28
HH:MM:SS:FF = frame number

From SMPTE Made Simple, Timeline Vista, Inc.
Synchronizing Diverse Sensors
Set Signals on a Common Timebase

Time (ms)
Kinect Capture System
Kinect Capture System

Dome Exterior

Dome Interior

Kinect v2

Clock System (LTC)

LTC Timecode

00:00
Kinect Capture System

- Kinect for Windows Adapter
- LTC Timecode

Dome Exterior
- Gigabyte Brix
- Capture Node
- Kinect v2

Dome Interior

Clock System (LTC)
00:00
Kinect Capture System

- Kinect v2
- Dome Exterior
- Dome Interior
- Capture Node
- Length Limited: USB 3.0
- Clock System (LTC) 00:00
- LTC Timecode
- Kinect for Windows Adapter
Kinect Capture System

- Kinect for Windows Adapter
- LTC Timecode

Dome Exterior

Dome Interior

Capture Node

Kinect v2

Clock System (LTC)
Kinect Capture System

- Kinect for Windows Adapter
- Ethernet
- LTC Timecode

Dome Exterior

Dome Interior

Capture Node

Kinect v2

Clock System (LTC)

00:00

Kinect Master

(Raw) 145 MB/s approximately 1.15 Gb/s
Dual NIC for realtime transfer
RGB+Depth Relative Timing

What Does Frame Start Mean?

![Diagram of RGB+Depth Relative Timing]
RGB+Depth Relative Timing

What Does Frame Start Mean?

Depth frame

Color frame

33.3ms

33.3ms
RGB+Depth Relative Timing

What Does Frame Start Mean?

Depth frame

Color frame

33.3ms

33.3ms

25 ms

Rolling shutter start

Time (ms)
Rolling Shutter Effects

Color

IR
Direction of motion

-12.5 ms

0 ms

+12.5 ms
Corner detection

- No compensation
+ Compensated
+ From Depth / IR

Direction of motion

-12.5 ms
0 ms
+12.5 ms
Fast Motion Artifacts

\[ v = \frac{f}{z} \cdot \frac{dy}{12.5 \text{ ms}} \]
Kinect Synchronization
Take Home Messages

- Off-the-shelf RGBD sensors are cheap, but you'll have to deal with synchronization & interference
- Modifying hardware is very impractical: PTP as an alternative
- Quick way to get 3D reconstruction
Kinect Calibration

Color

IR

Depth
... Back To The Checkerboard
YOU MISSED A SPOT
Kinect Calibration
Take Home Messages

- Combining RGB+IR+Depth will require specialized calibration
- Static checkerboard to calibrate intrinsics and extrinsics between color & depth per Kinect, SfM between sensors
- Multiple RGB-D sensors may produce additional noise and interferences
Break/Questions

After the break:

DRZ

Thabo Beeler

Derek Bradley
Backup Slides
Panoptic Studio Architecture
Time Alignment Across Subsystems

Clock System

Master Node

- Local Node
- 24 VGA Cameras
- 480 VGA Cameras (20 Panels)
- 10 Kinects

NVIDIA Optic Fiber
GbE
Genlock
HD-SDI
LTC
VGA Video
USB 3.0
Time Align.
Panoptic Studio Architecture
Time Alignment Across Subsystems

Clock System

Master Node

Local Node

24 VGA Cameras

480 VGA Cameras (20 Panels)

24 VGA Cameras

480 VGA Cameras (20 Panels)

31 HD Cameras

(31 Panels)

10 Kinects

10 Kinects

5 DLP Projectors

(5 Panels)

Optic Fiber  GbE  Genlock  HD-SDI  LTC  VGA Video  USB 3.0  Time Align.