TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Performance and Accuracy Assessment of Nvidia's Omniverse Isaac Sim for Generating Synthetic Data from Real-world Scenarios

Bachelorthesis

Patrick Noras

07. December 2023

Outline

1. Motivation
2. Synthetic Data Generation
3. Methodology
4. Results
5. Conclusion

1. Motivation

- Task: Train Autonomous Vehicles

1. Motivation

- Task: Train Autonomous Vehicles
  - Traditional Approach:
    1. Acquire a Vehicle

Image Source: https://ev-database.org/de/pkw/1941/Opel-Corsa-Electric-50-kWh

1. Motivation

- Task: Train Autonomous Vehicles
  - Traditional Approach:
    1. Acquire a Vehicle
    2. Acquire Sensors

Image Source: https://www.stereolabs.com/products/zed-2

Image Source: https://www.robosense.ai/en/rslidar/RS-Helios

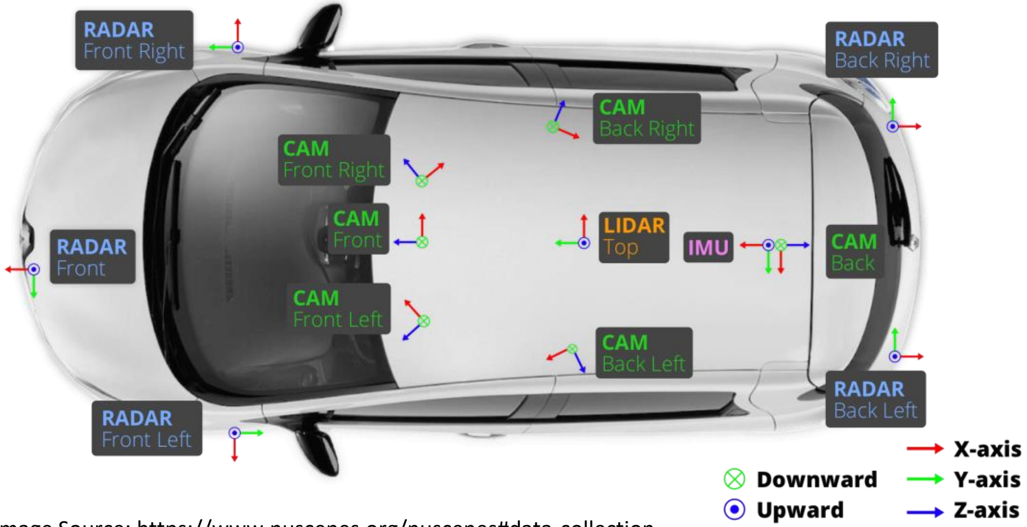


Image Source: https://www.nuscenes.org/nuscenes#data-collection

1. Motivation

- Task: Train Autonomous Vehicles
  - Traditional Approach:
    1. Acquire a Vehicle
    2. Acquire Sensors
    3. Collect data

Image Source: https://www.latimes.com/business/story/2023-10-11/waymo-driverless-taxi-launch-in-santa-monica-met-with-excitement-tension

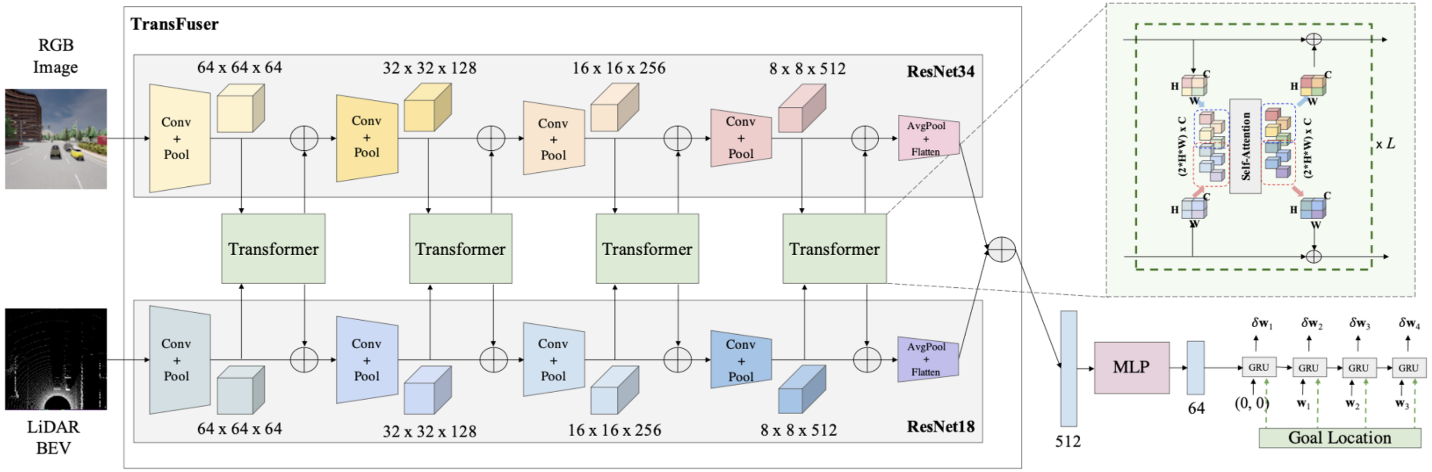
1. Motivation

- Task: Train Autonomous Vehicles
  - Traditional Approach:
    1. Acquire a Vehicle
    2. Acquire Sensors
    3. Collect data
    4. Label data

Image Source: https://medium.com/intro-to-artificial-intelligence/semantic-segmentation-udaitys-self-driving-car-engineer-nanodegree-c01eb6eaf9d

1. Motivation

- Task: Train Autonomous Vehicles
  - Traditional Approach:
    1. Acquire a Vehicle
    2. Acquire Sensors
    3. Collect data
    4. Label data
    5. Train Network



Aditya Prakash et al. „Multi-Modal Fusion Transformer for End-to-End Autonomous Driving“ CVPR, 2021.

What are the Problems?

- Equipment is very expensive!
  - LiDAR: 1.200€ - 12.000€
  - Stereo Camera: > 200€
  - Car: < 40.000€
- High time investment
  - Collecting data
  - Labelling data
- Bias in dataset
- Exhausting other Resources…

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

What are the Problems?

- Equipment is very expensive!
  - LiDAR: 1.200€ - 12.000€
  - Stereo Camera: > 200€
  - Car: < 40.000€
- High time investment
  - Collecting data
  - Labelling data
- Bias in dataset
- Exhausting other Resources…

➡ Generate Synthetic Data

2. Synthetic Data Generation

- What is Synthetic Data?
  - Data that has been created artificially
  - Used as training data for transfer learning
  - Should cover phenomenon of real data

➡ Capture realism as much as possible

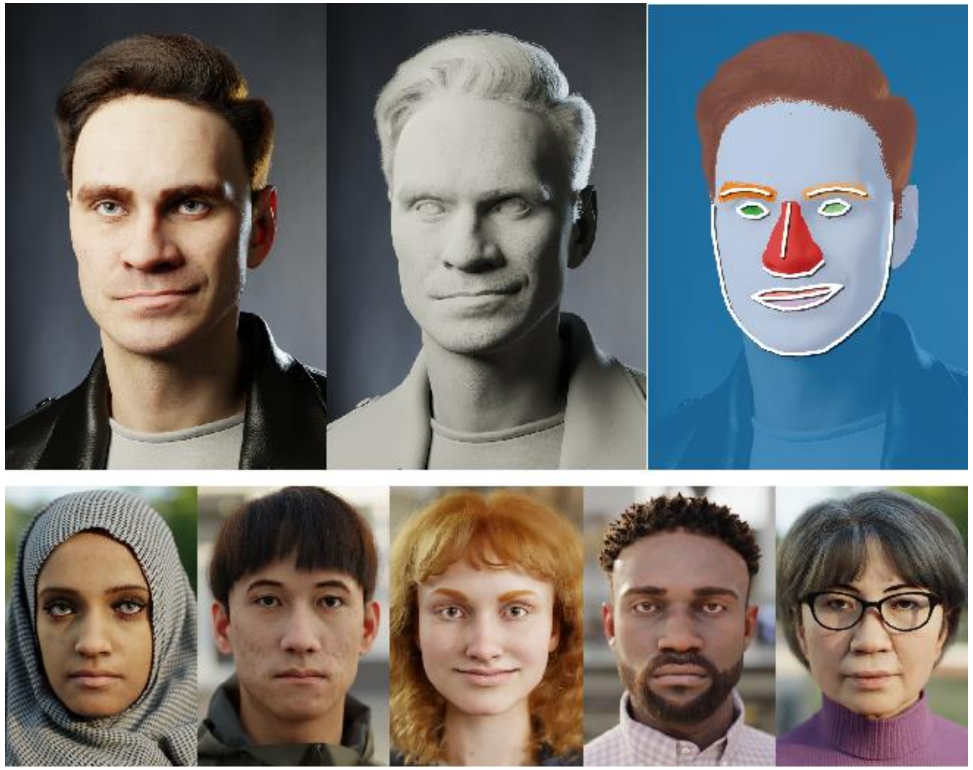
Erroll Wood et al. “Fake It Till You Make It: Face analysis in the wild using synthetic data alone”

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

2. Synthetic Data Generation

- What is Synthetic Data?
  - Data that has been created artificially
  - Used as training data for transfer learning
  - Should cover phenomenon of real data

  ➡ Capture realism as much as possible
- What are the Challenges?
  - Expense of computational resources and time
  - Task dependent
  - Domain gap problem



Erroll Wood et al. “Fake It Till You Make It: Face analysis in the wild using synthetic data alone”

Domain Gap

- Content Gap
  - Variety of data
  - Approximate p_{real} such that $p_{syn} \approx p_{real}$

➡ Domain randomization

Domain Gap

- Content Gap
  - Variety of data
  - Approximate p_{real} such that $p_{syn} \approx p_{real}$

  ➡ Domain randomization
- Appearance Gap
  - Materials, Assets
  - Rendering systems

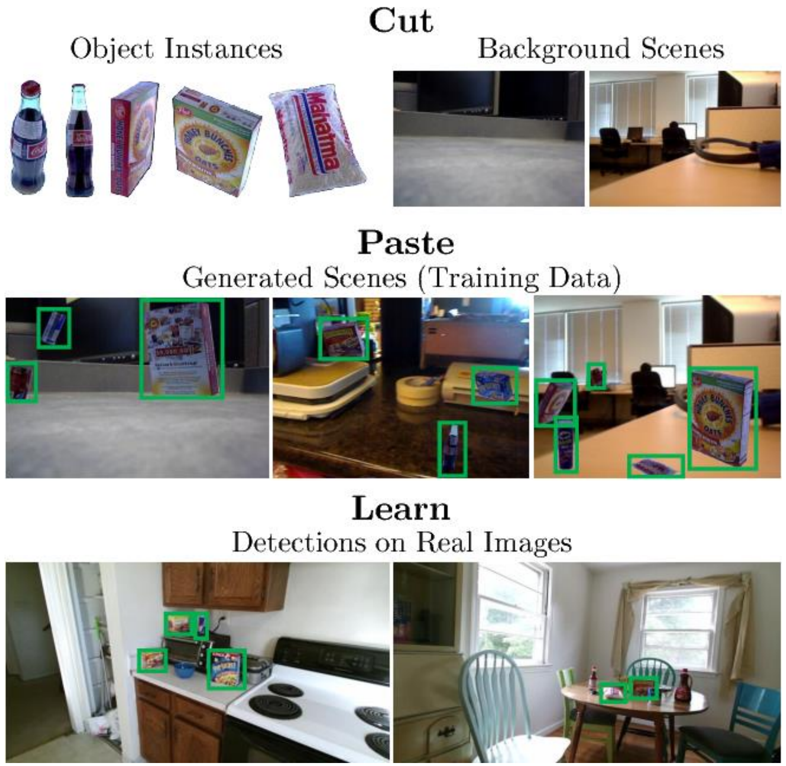
  ➡ Sophisticated Generation Methods (Task dependent)

Domain Gap

- Content Gap
  - Variety of data
  - Approximate p_{real} such that $p_{syn} \approx p_{real}$

  ➡ Domain randomization
- Appearance Gap
  - Materials, Assets
  - Rendering systems

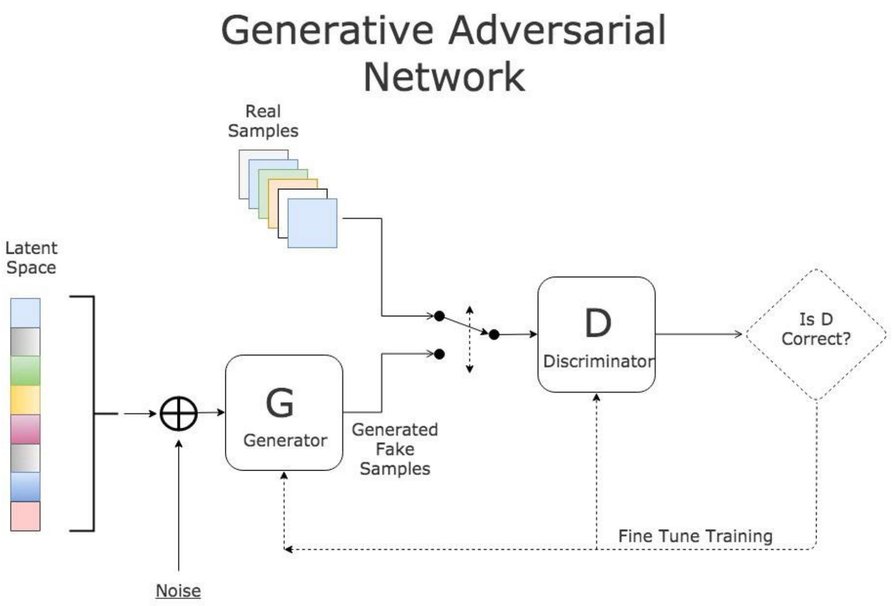
  ➡ Sophisticated Generation Methods (Task dependent)
- What are Generation Methods?
  - Simple Augmentations



Dwibedi et al."Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection"

Domain Gap

- Content Gap
  - Variety of data
  - Approximate p_{real} such that $p_{syn} \approx p_{real}$
  - ➡ Domain randomization
- Appearance Gap
  - Materials, Assets
  - Rendering systems
  - ➡ Sophisticated Generation Methods (Task dependent)
- What are Generation Methods?
  - Simple Augmentations
  - Generative Models, GANs



Goodfellow, Ian, et al. "Generative adversarial networks."

Domain Gap

- Content Gap
  - Variety of data
  - Approximate p_{real} such that $p_{syn} \approx p_{real}$

  ➡ Domain randomization
- Appearance Gap
  - Materials, Assets
  - Rendering systems

  ➡ Sophisticated Generation Methods (Task dependent)
- What are Generation Methods?
  - Simple Augmentations
  - Generative Models, GANs
  - <u>Simulators</u>

Domain Gap

- Content Gap
  - Variety of data
  - Approximate p_{real} such that $p_{syn} \approx p_{real}$

  ➡ Domain randomization
- Appearance Gap
  - Materials, Assets
  - Rendering systems

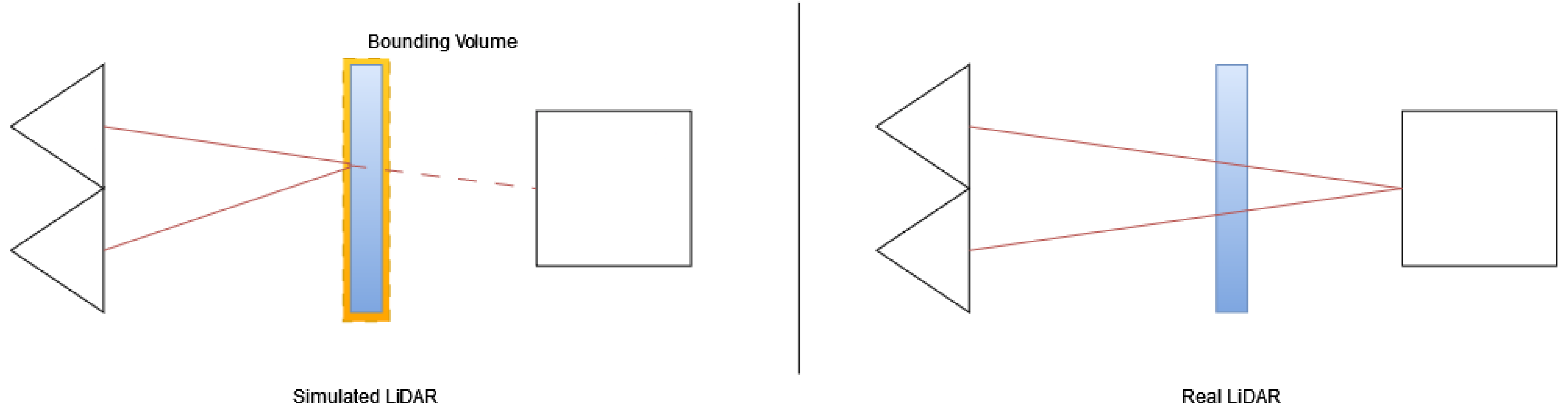
  ➡ Sophisticated Generation Methods (Task dependent)
- What are Generation Methods?
  - Simple Augmentations
  - Generative Models, GANs
  - Simulators
  - Many more…

How to achieve Realism in Simulators?

- High-quality Assets and Materials
- Accurate implementation of Sensors
  - Cameras (Synthetic Images)
    - State-of-the-art Rendering
  - LiDAR (Synthetic Point Clouds)
    - Accurate ToF implementation

➡ Implementations vary highly across Simulators

Example: Naïve LiDAR ToF implementation with Glass

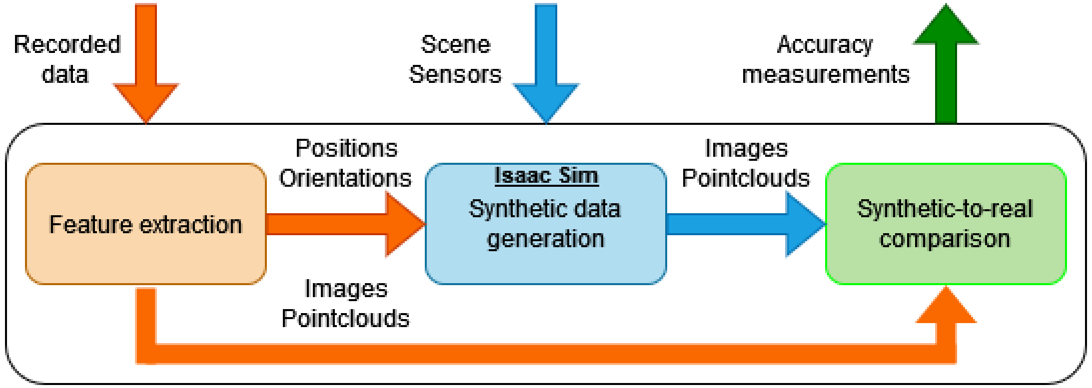


3. Methodology

- What are the Goals?
  - Creation of an accurate Digital Twin
  - Replicate real Scenario
  - Audit Isaac Sim’s proficiency in data synthesis

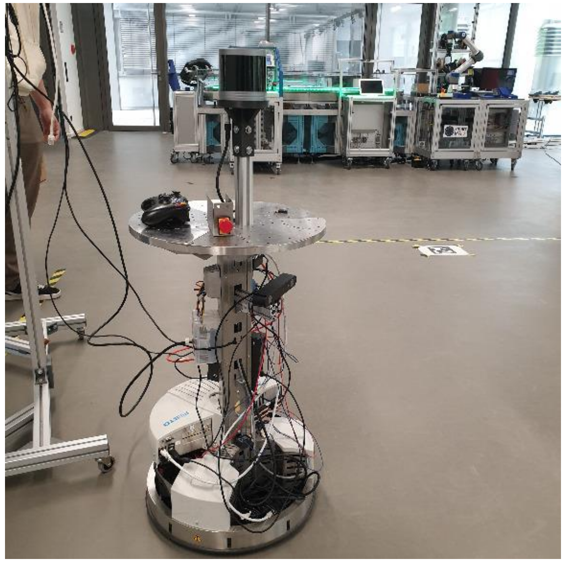
3. Methodology

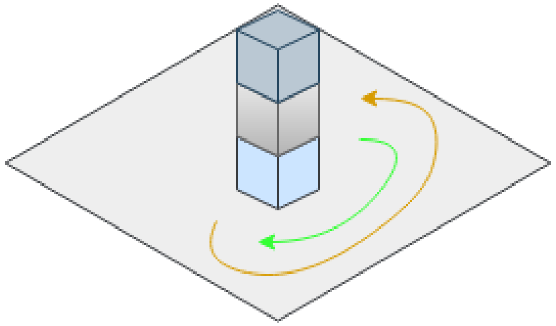
- What are the Goals?
  - Creation of an accurate Digital Twin
  - Replicate real Scenario
  - Audit Isaac Sim's proficiency in data synthesis
- How do we assess the Synthetic Data Generation?
  - Evaluate synthetic data against real counterpart
  - Cover different test cases (Diffuse, transparent and highly reflective Objects)

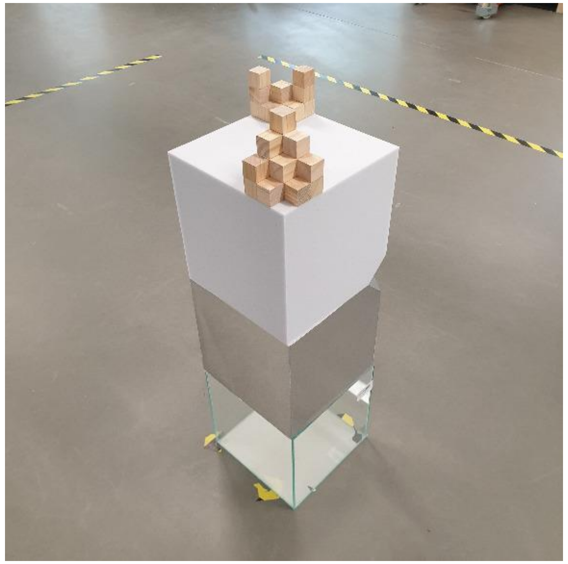


Physical Setup

- Sensors
  - ZED 2 Stereo Camera
  - LiDAR RS-Helios-32-5515
- Environment
  - Cube Tower with different Materials
  - Controlled Environment
- Scan
  - Multiple rounds around the Cube Tower
  - RGB Images, Point Clouds, Pose and Timestamp data







Data Processing

- Inconsistent Amount of data collected
  - No Pose for every recorded Image and Point Cloud

Name	Amount
Poses	3281
Left Images	3332
Point Clouds	1161

Data Processing

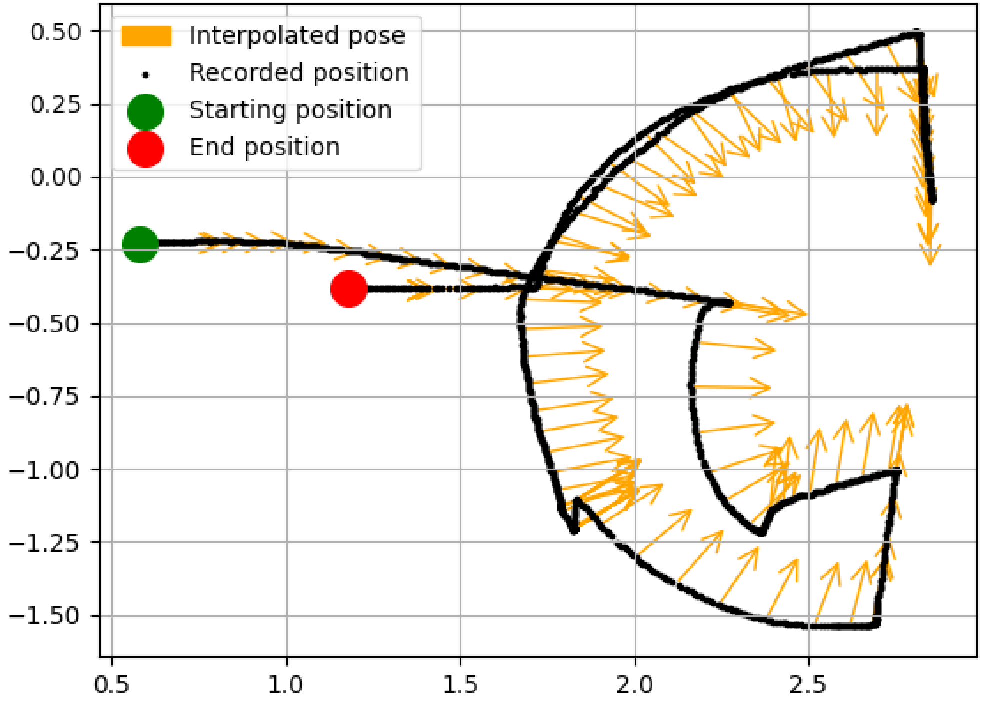
- Inconsistent Amount of data collected
  - No Pose for every recorded Image and Point Cloud
- Interpolate missing Pose information

Name	Amount
Poses	3281
Left Images	3332
Point Clouds	1161

Algorithm 1

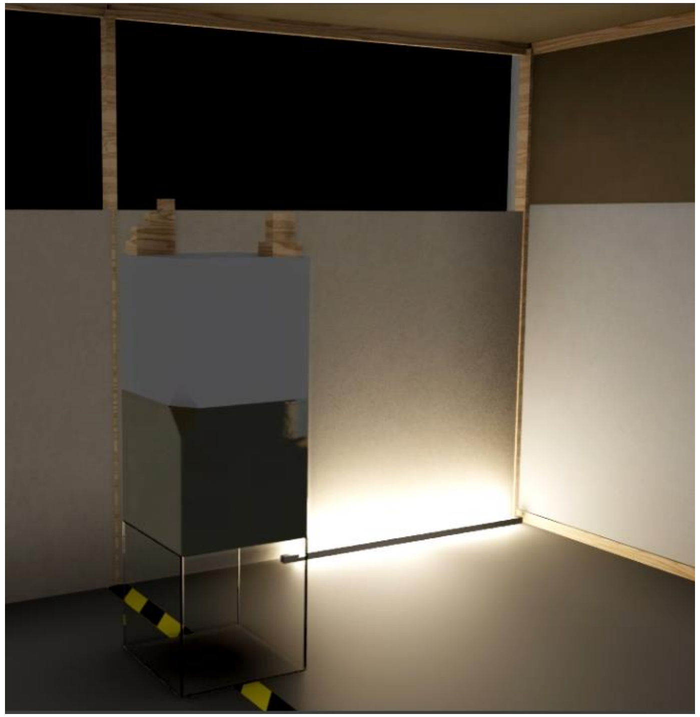
Input: Sensor tuple $(\tau_\lambda, d_\lambda) \in S$ where τ_λ is the timestamp at index λ and path data P
Output: $(\tau_\lambda, (p_k, o_k))$, where (p_k, o_k) is the interpolated pose

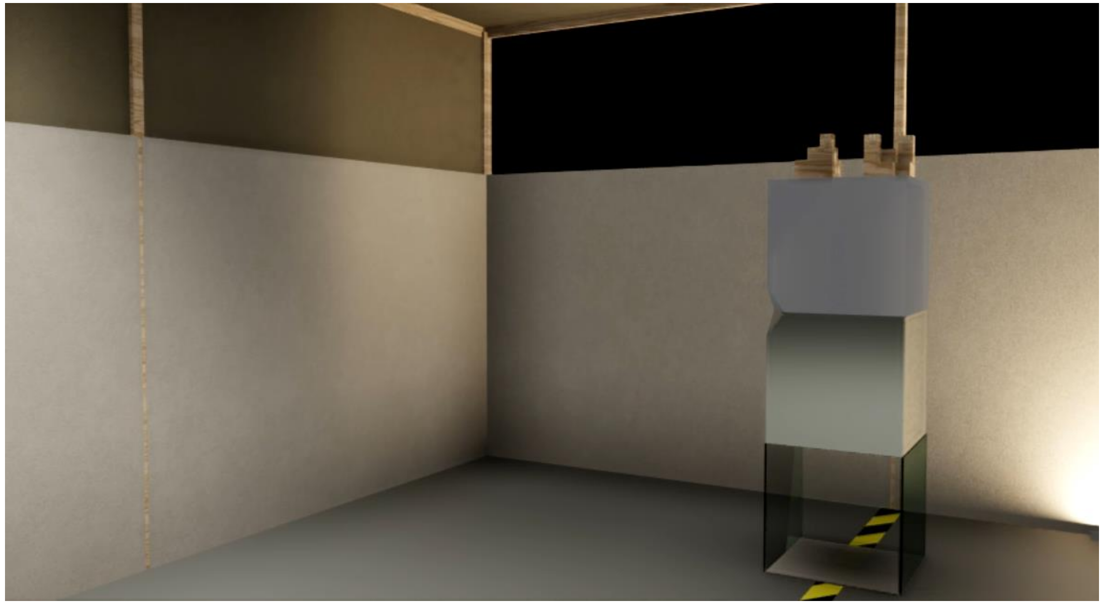
function INTERPOLATEPOSE$(\tau_\lambda, d_\lambda, P)$
  $k = 1, t_{k-1} = P[k-1], t_{k+1} = P[k+1]$
  while $\neg(t_{k-1} \leq \tau_\lambda < t_{k+1}) \wedge k \leq |P|$ **do**
    $k = k + 1$
    $t_{k-1} = P[k-1], t_{k+1} = P[k+1]$
  end while
  if $k > |P|$ **then**
    return Out of bounds
  end if
  if $\tau_\lambda == t_{k-1}$ **then**
    return $(\tau_\lambda, (p_{k-1}, o_{k-1}))$
  end if
  $a = \frac{\tau_\lambda - t_{k-1}}{t_{k+1} - t_{k-1}}$
  - Do linear interpolation for the position
  $p_k = (1-a)p_{k-1} + a p_{k+1} +$
  - Do SLERP for the orientation
  $\theta = \arccos\left(\frac{o_{k-1} \cdot o_{k+1}}{\|o_{k-1}\| \|o_{k+1}\|}\right)$
  $o_k = \frac{\sin((1-a)\theta) o_{k-1} + \sin(a\theta) o_{k+1}}{\sin(\theta)}$
  return $(\tau_\lambda, (p_k, o_k))$
end function



Virtual Setup

- Real Environment has been modelled within Isaac Sim



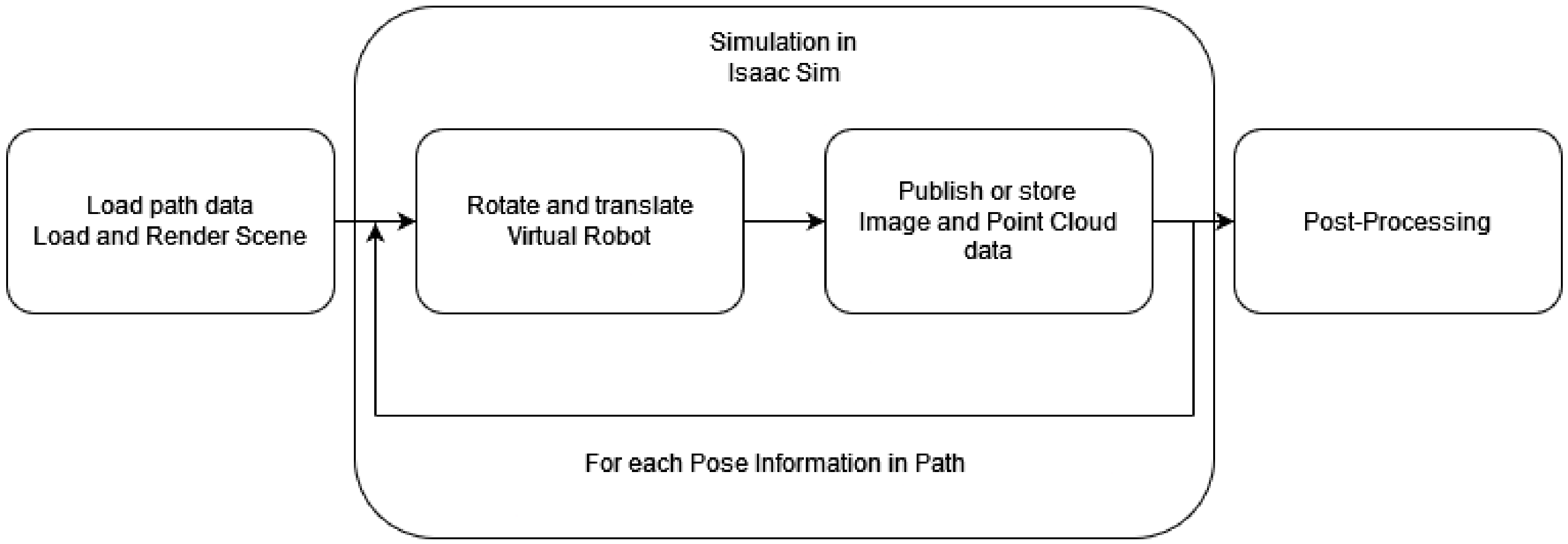


Virtual Sensors

- Camera
  - Standard stage camera
- LiDAR
  - PhysX LiDAR (Naïve ToF implementation)
  - RTX LiDAR (Considers predefined Material characteristics)
- Isaac Sim's coordinate convention
  - World: Z = Up, X = Forward
  - Camera: Y = Up, -Z = Forward

➡ Rotate Sensors by -90° around the Z and Y axis, such that they look at the virtual Cube Tower

Simulation and Data Synthesis



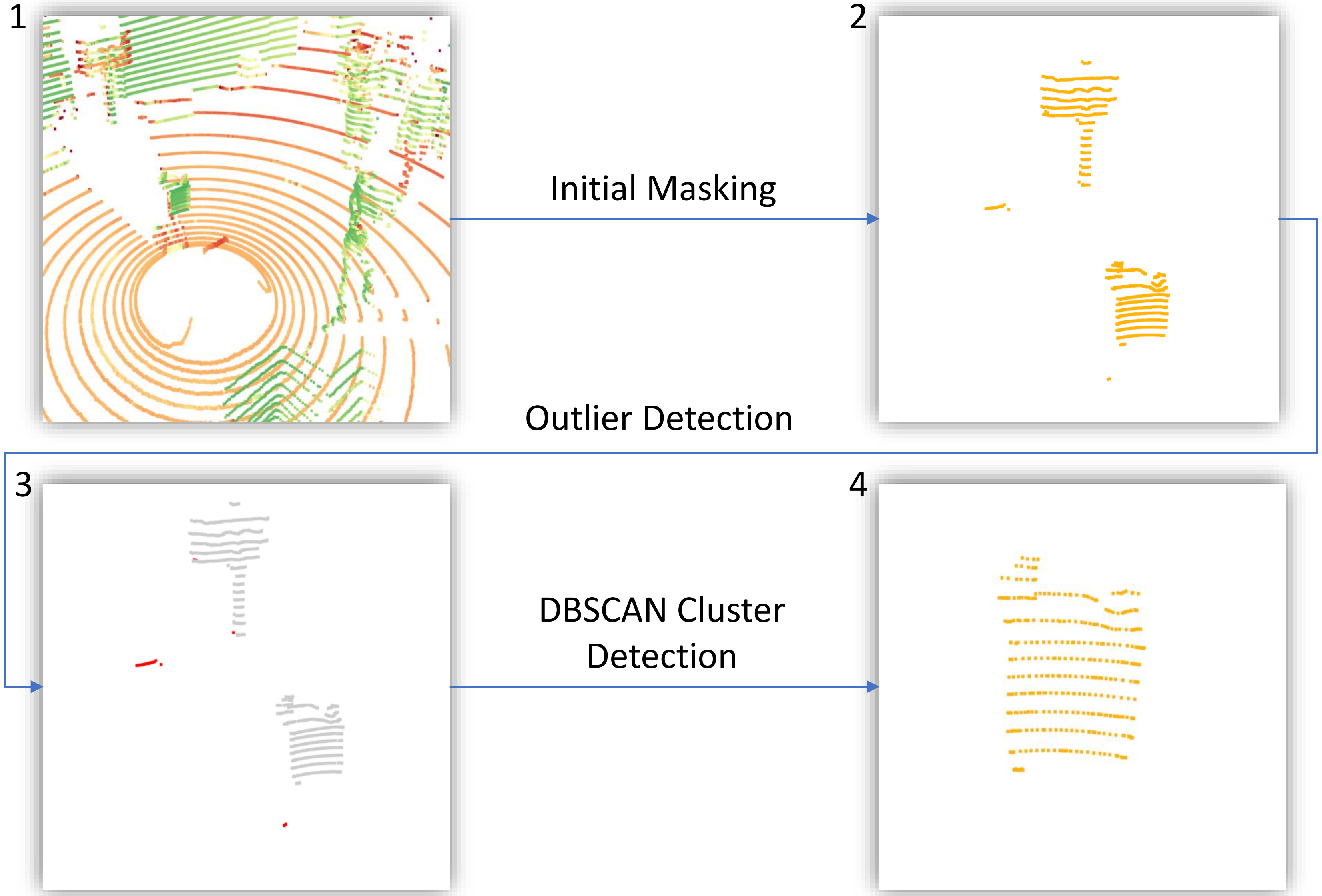
Post-Processing

- Synthetic Images
  - Virtual Cameras are not affected by noise

  ➡ Apply Gaussian Noise filter
- Synthetic Point Clouds
  - We are only interested in Point Cloud data of our Cube Tower
  - Virtual and Real LiDAR have a 360° FoV

  ➡ Unnecessary point data for structures we are not interested in

Post-Processing



4. Results

- Image Similarity Metrics

Metric	Description	Value Range
RMSE	Root-mean-square error	[0, ∞)
PSNR	Peak signal-to-noise ratio	[0, ∞)
SRE	Signal-to-reconstruction error ratio	[0, ∞)
SSIM	Structural similarity index	[0, 1]

- Point Cloud Similarity Metrics

Metric	Description	Value Range
RMSE	Root-mean-square error	[0, ∞)
HD	Hausdorff distance	[0, ∞)
CD	Chamfer distance	[0, ∞)

Image Quality Measurements

Metric	Average Value		Best Value	
	Interpolated Path	Non-interpolated Path	Interpolated Path	Non-interpolated Path
RMSE	0.18	0.18	0.13	0.13
PSNR	15.04	15.04	17.23	17.23
SRE	51.52	51.52	52.80	52.80
SSIM	0.71	0.71	0.78	0.78

- Path interpolation had no significant change on Synthetic Images
- Satisfying average SSIM and SRE values

Image Quality Measurements

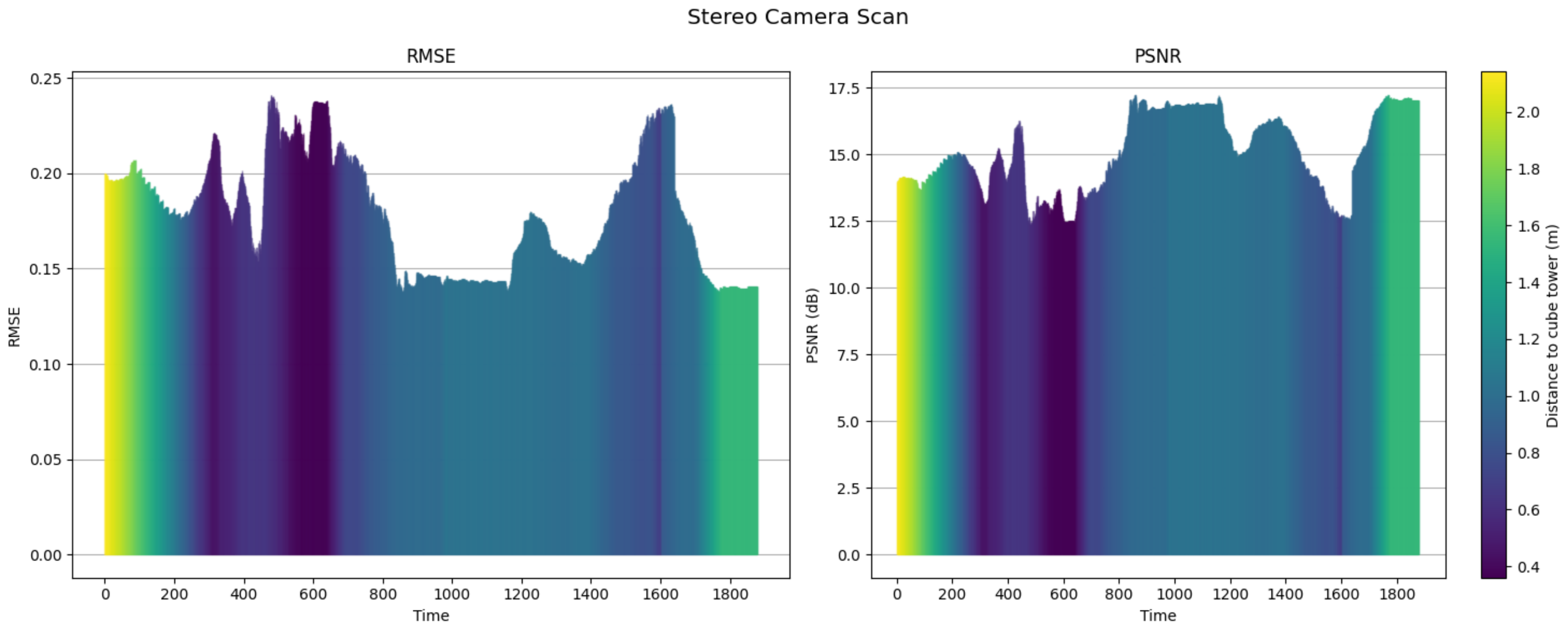


Image Quality Measurements

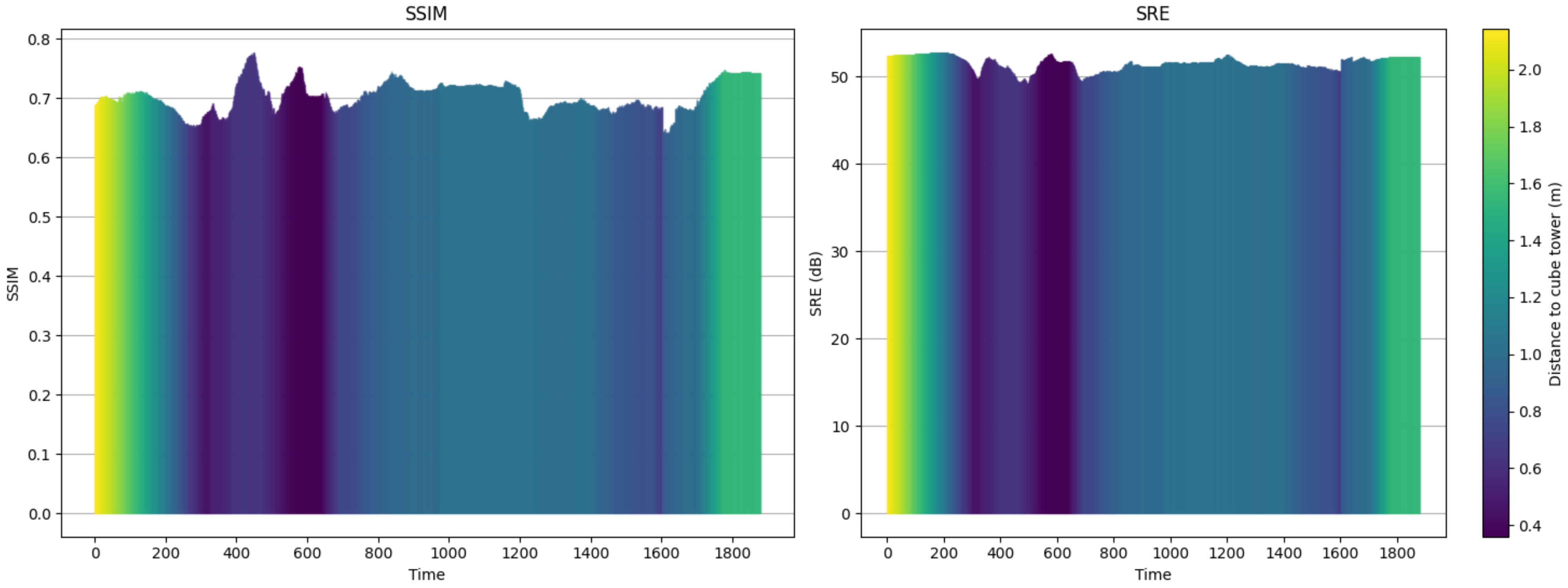
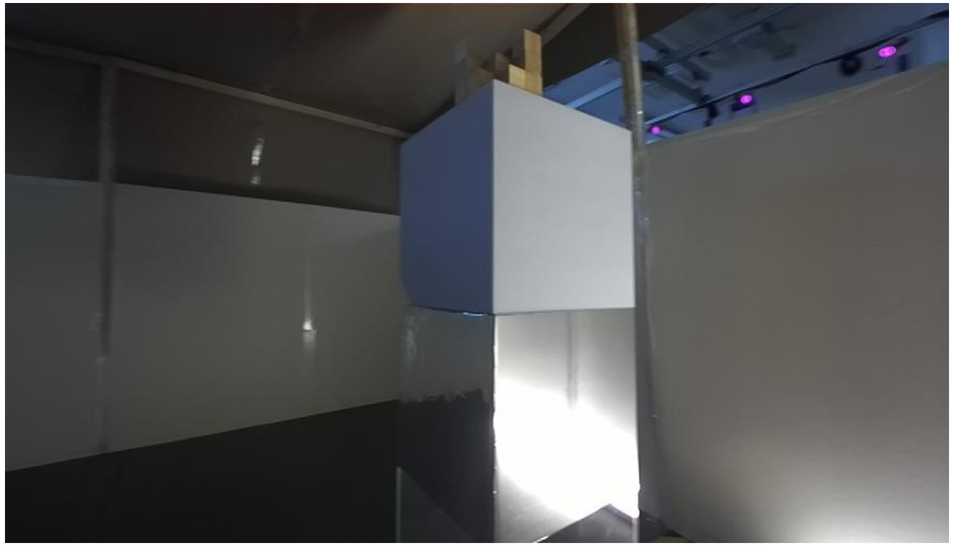
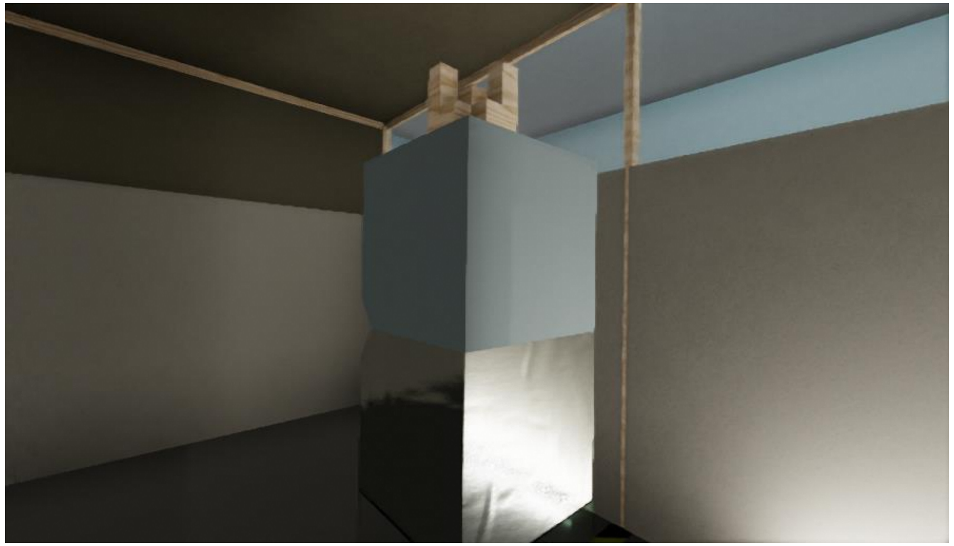


Image Quality Measurements

Best
SSIM





Best
SRE

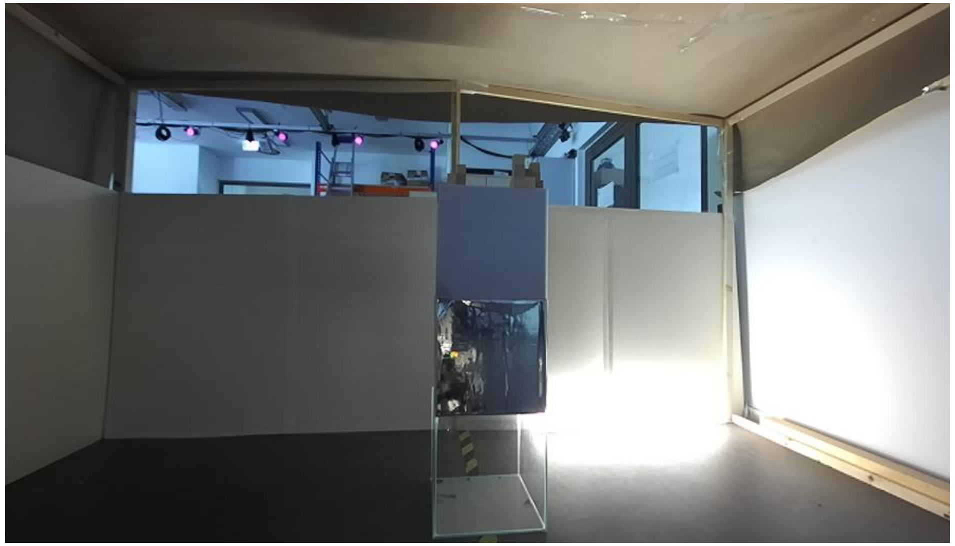
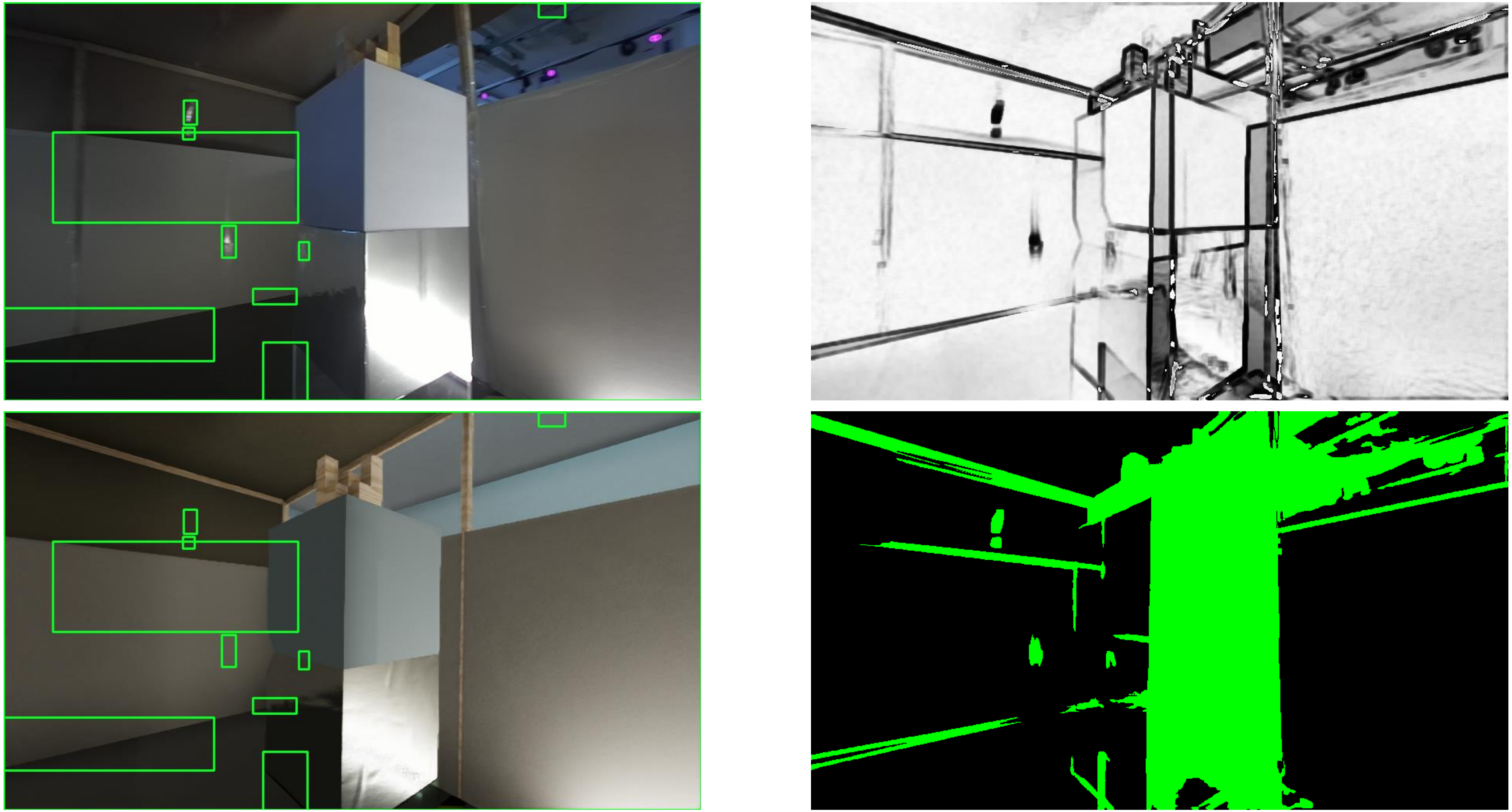


Image Quality Measurements

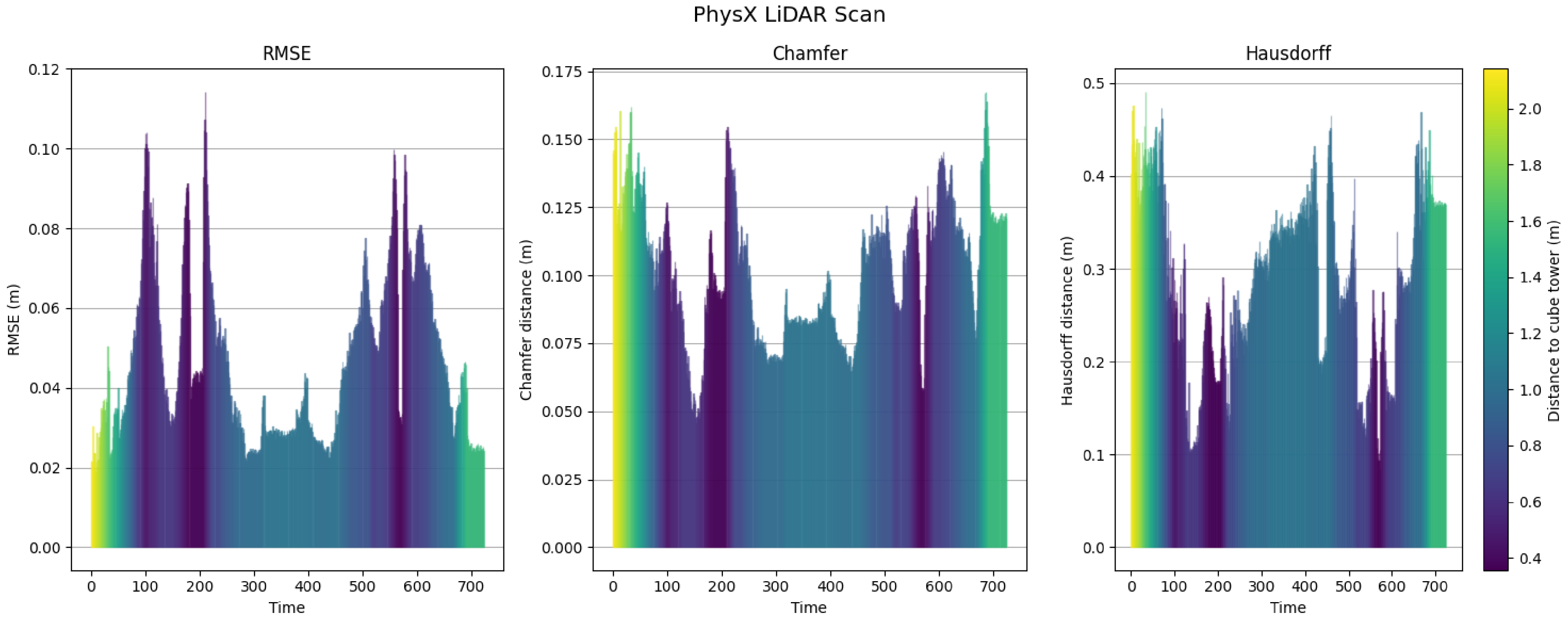


Point Cloud Similarity Measurements

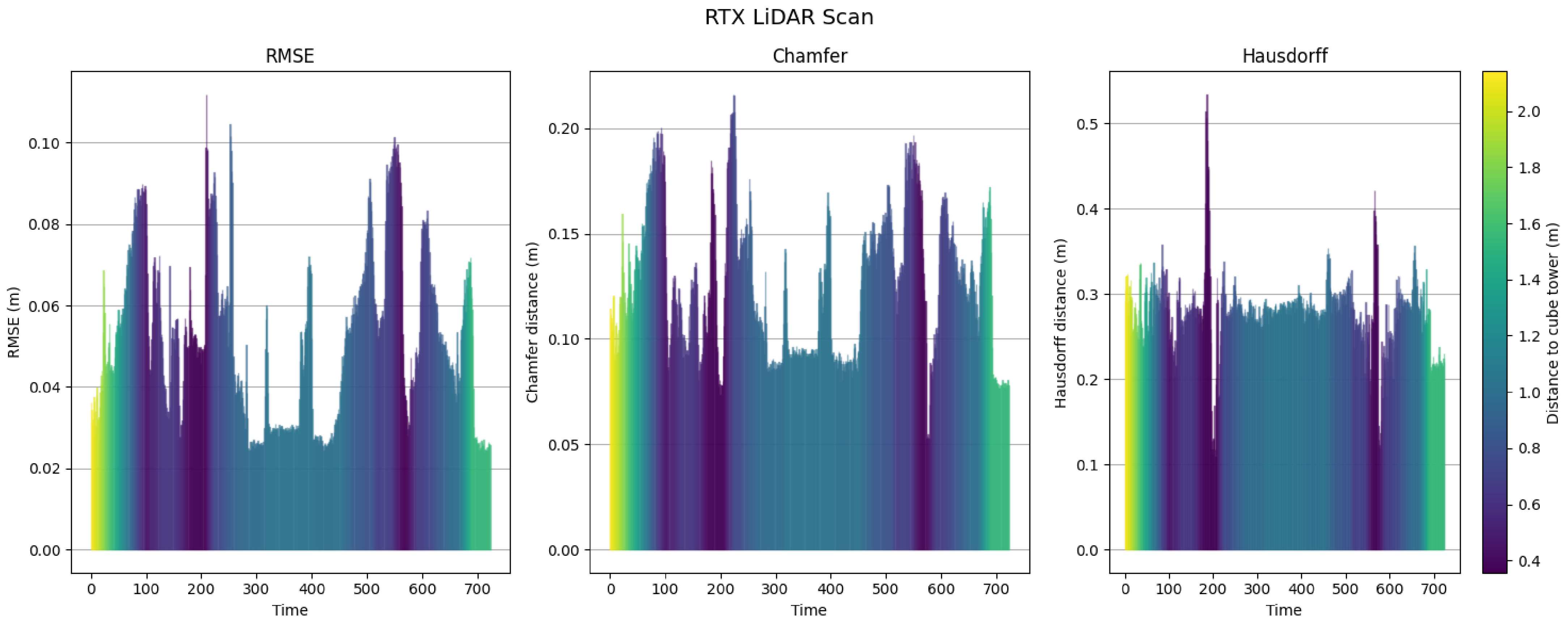
Metric	Average Value			
	Interpolated Path RTX	Non-interpolated Path RTX	Interpolated Path PhysX	Non-interpolated Path PhysX
RMSE	0.051	0.054	**0.046**	0.047
CD	0.12	0.13	**0.1**	**0.1**
HD	**0.27**	0.28	0.28	0.28

- Path interpolation yields in general better results
- Best RMSE and CD measures for PhysX LiDAR
- However best HD measure for RTX LiDAR

Point Cloud Similarity Measurements

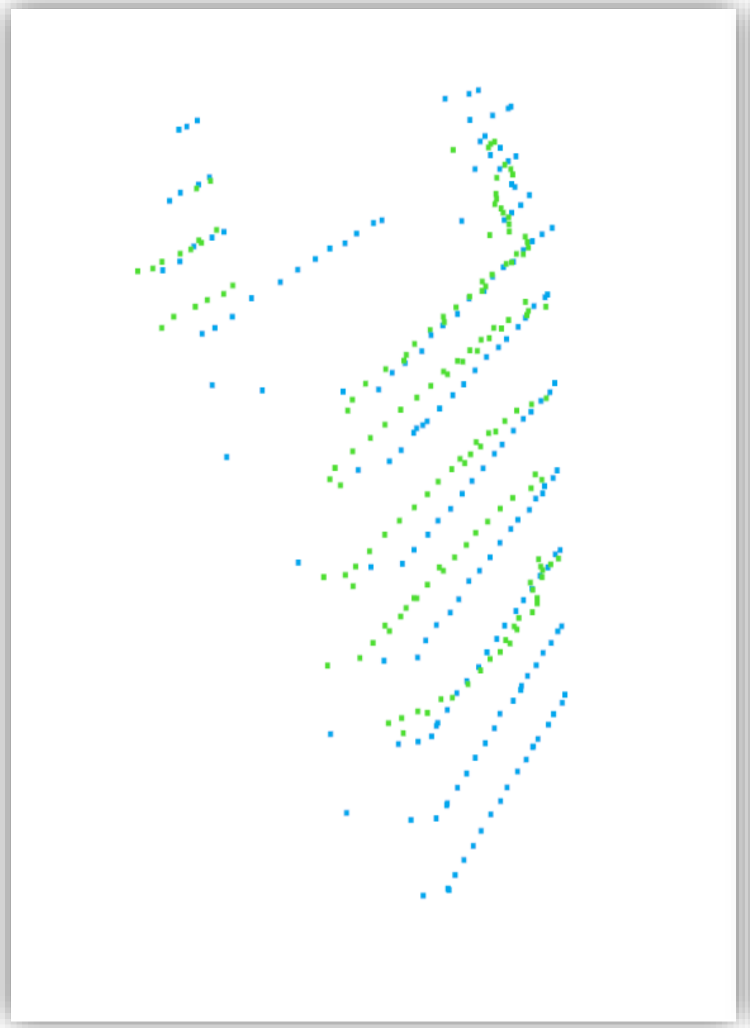


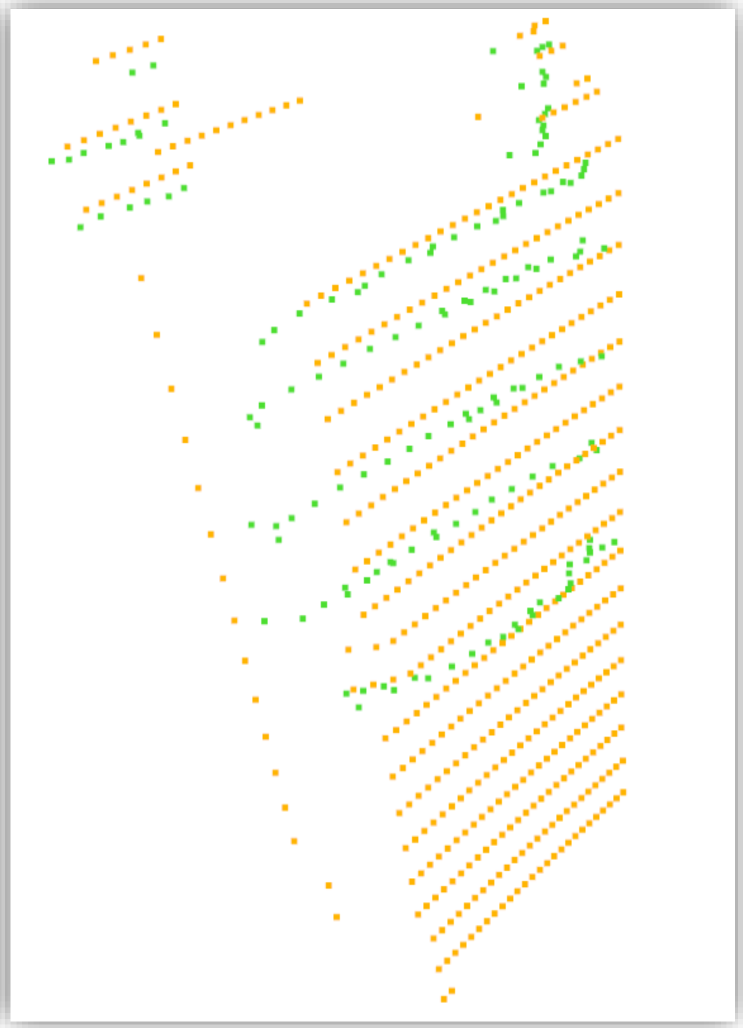
Point Cloud Similarity Measurements



Point Cloud Similarity Measurements

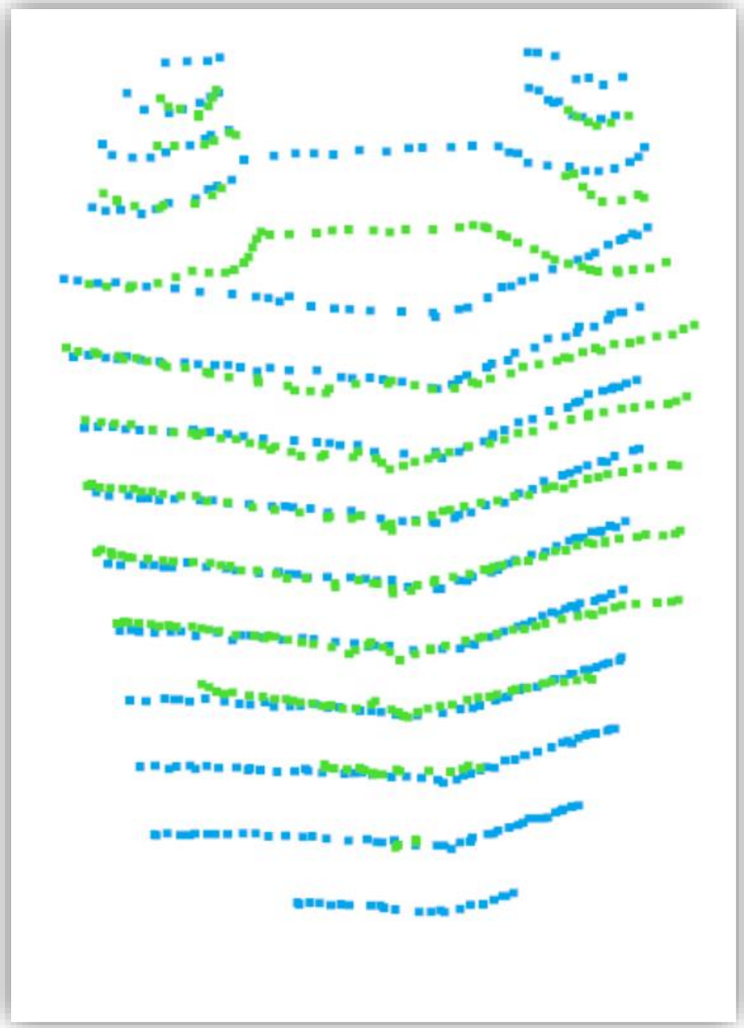
- Green = Real LiDAR Scan
  Blue = RTX LiDAR Scan
  Orange = PhysX LiDAR Scan
- RTX and PhysX LiDAR detect Mirroring Cube
- Only PhysX LiDAR detects Glass Cube

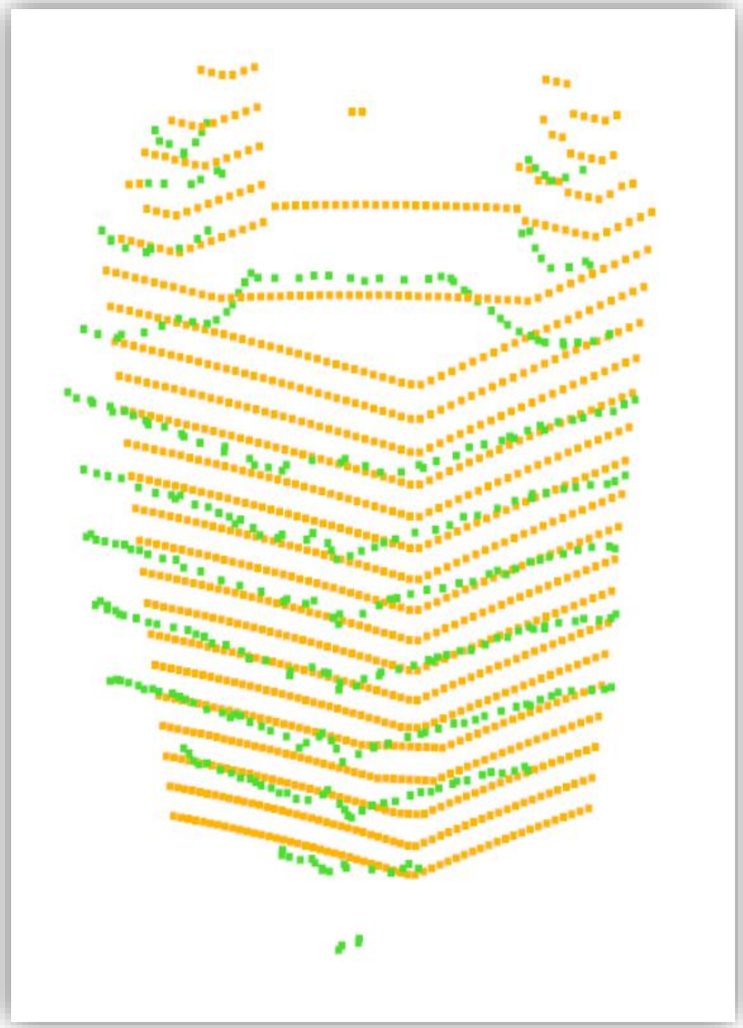




Point Cloud Similarity Measurements

- PhysX LiDAR exhibits too much detail
- RTX LiDAR introduces noise into the data





5. Conclusion

- Synthetic Data Generation is an important tool for today's Industry

5. Conclusion

- Synthetic Data Generation is an important tool for today’s Industry
- Closing the Domain Gap with Simulators
  - Need of high quality Assets
  - State-of-the-art Rendering
  - Accurate implementation of Sensors

→ Common LiDAR ToF implementation pitfalls

5. Conclusion

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

- Synthetic Data Generation is an important tool for today's Industry
- Closing the Domain Gap with Simulators
  - Need of high quality Assets
  - State-of-the-art Rendering
  - Accurate implementation of Sensors

  → Common LiDAR ToF implementation pitfalls
- Isaac Sim
  - Satisfying Synthetic Images with just an approximated Model of Real Environment
  - PhysX LiDAR follows common Naïve LiDAR implementation
  - RTX LiDAR introduces a better attempt at simulating real LiDAR sensors

What could have been improved?

- Improved controlled Environment for Scans

What could have been improved?

- Improved controlled Environment for Scans
- More accurate Assets of real World Scenario

What could have been improved?

- Improved controlled Environment for Scans
- More accurate Assets of real World Scenario
- Well-known Start and End Positions with exact Distance measurements

What could have been improved?

- Improved controlled Environment for Scans
- More accurate Assets of real World Scenario
- Well-known Start and End Positions with exact Distance measurements
- Comparing data synthesis capabilities of other Robotics Simulators in contrast to Isaac Sim

What could have been improved?

- Improved controlled Environment for Scans
- More accurate Assets of real World Scenario
- Well-known Start and End Positions with exact Distance measurements
- Comparing data synthesis capabilities of other Robotics Simulators in contrast to Isaac Sim
- Assessment of Isaac Sim’s depth sensing capabilities

What could have been improved?

- Improved controlled Environment for Scans
- More accurate Assets of real World Scenario
- Well-known Start and End Positions with exact Distance measurements
- Comparing data synthesis capabilities of other Robotics Simulators in contrast to Isaac Sim
- Assessment of Isaac Sim’s depth sensing capabilities
- Applying ML model to determine performance by training with synthetic and real data
    - Object recognition, Segmentation, 3D Pose Estimation

What could have been improved?

- Improved controlled Environment for Scans
- More accurate Assets of real World Scenario
- Well-known Start and End Positions with exact Distance measurements
- Comparing data synthesis capabilities of other Robotics Simulators in contrast to Isaac Sim
- Assessment of Isaac Sim’s depth sensing capabilities
- Applying ML model to determine performance by training with synthetic and real data
  - Object recognition, Segmentation, 3D Pose Estimation
- Hybrid Solutions
  - Simulators with GANs along with domain randomization

Thank you for your attention!