

D5.2 - Final Test Report

Project number:	248623
Project acronym:	TACO
Project title:	Three-dimensional Adaptive Camera with Object Detection and Foveation
Start date of the project:	February 1 st , 2010
Duration:	42 months

Deliverable type:	REPORT
Deliverable reference number:	248623 / D5.2 / FINAL V1.1
Deliverable title:	Final Test Report
WP contributing to the deliverable:	WP5
Due date:	July 2013 – M42
Actual submission date:	July 31 st 2013 (V1.0), September 12 th 2013 (V1.1)

Responsible organisation:	TUW
Authors:	TUW (M. Vincze, A. Aldoma), OTL (P Murcutt, R King), Shadow Robot (T. Oliver, M. Addison), SINTEF (Trine Kirkhus, Jens Thielemann, Petter Risholm)
Abstract:	This document contains the final tests performed on the TACO sensor as well as the experiences from end-users with it.
Keywords:	Test, Use Cases Report

Dissemination level:	Public
Revision:	FINAL V1.1

Instrument:	STREP
Thematic Priority:	ICT

Table of Contents

- List of Figures 4
- List of Tables 6
- 1 Executive Summary 7
- 2 System test and verification 8
 - 2.1 Summary of system integration – hardware/software, system test and verification..... 8
 - 2.2 Distance measurements..... 9
 - 2.3 Observed Sensor Data Artifacts11
 - 2.4 Changing mirror plans12
 - 2.4.1 Example of 1Hz mirror plan.....15
 - 2.4.2 Example using 10Hz linear mirror plan.....16
 - 2.4.3 Example using 10Hz foveating mirror plans.....17
 - 2.4.4 Combination of 10Hz linear and 1Hz foveated trajectories.....18
 - 2.5 Attention algorithms used for foveation20
 - 2.5.1 Range model (' attentionConfig_RangeModel')20
 - 2.5.2 Contour tracker ('attentionConfig_Tracking')20
 - 2.5.3 Distance based foveation (' attentionConfig_Distance')21
 - 2.5.4 Other foveation methods21
 - 2.5.5 Sending mirror plans over the network from robot to the TACO sensor.....21
 - 2.6 Influence of external illumination in the 1550nm band.....22
- 3 Use Cases Evaluation.....24
 - 3.1 3D Inspection & Augmented Reality (OTL)24
 - 3.1.1 3D Inspection.....24
 - 3.1.1.1 Set-up24
 - 3.1.1.2 Technical description of the method24
 - 3.1.1.3 Evaluation method.....25
 - 3.1.1.3.1 Feature Detection25
 - 3.1.1.3.2 Averaging of point cloud data.....26
 - 3.1.1.4 Results26
 - 3.1.1.4.1 Feature Detection26
 - 3.1.1.4.2 Averaging of point cloud data.....31
 - 3.1.1.5 Discussion.....34
 - 3.1.1.5.1 Sensor warm-up34
 - 3.1.1.5.2 Missing points & intensity data35
 - 3.1.1.5.3 Non-Reflective Surfaces36
 - 3.1.1.5.4 X Axis oscillation.....37
 - 3.1.1.5.5 Foveation38
 - 3.1.1.5.6 Conclusions38
 - 3.1.2 Augmented Reality39
 - 3.1.2.1 Set-up39
 - 3.1.2.2 Technical description of the method39
 - 3.1.2.3 Evaluation method.....40
 - 3.1.2.4 Results40
 - 3.1.2.4.1 Back drop40
 - 3.1.2.4.2 Marker selection40
 - 3.1.2.4.3 Marker detection42
 - 3.1.2.5 Discussion.....43
 - 3.2 Public Safety (SHADOW).....43
 - 3.2.1 Visual serving.....44
 - 3.2.1.1 Set-up44
 - 3.2.1.2 Technical description of the method44
 - 3.2.1.3 Evaluation method.....45
 - 3.2.1.3.1 TACO Sensor Density change48
 - 3.2.1.4 Results49
 - 3.2.1.5 Discussion.....49

3.3	Home Grasping (TUW).....	50
3.3.1	Self-localisation	50
3.3.2	Object detection.....	50
3.3.2.1	Set-up	50
3.3.2.2	Artifacts with the TACO sensor	51
3.3.2.3	Technical description of the method	52
3.3.2.4	Results	53
3.3.2.5	Effect of direct sunlight (Kinect vs TACO).....	56
3.3.2.6	Discussion.....	58
4	Does Foveation Make Sense?	59
4.1	Summary from the uses cases regarding foveation and discussion	59
5	Application map and lessons learned – the TACO sensor.....	61
5.1	Foveation and foveating systems.....	61
5.2	Attention algorithms	62
5.3	Real-time mirror control.....	63
5.4	Foveation Hardware	63
5.5	Applications for the developed hardware.....	64
5.6	Applications for the developed MEMS scanner technology	64
5.7	Applications for alternative hardware	65
5.8	Exploitation.....	65
5.9	General lessons learned	65
6	References.....	66
7	Appendix	67
7.1	Public Safety Use Case Data.....	67
7.1.1	TACO Sensor, Normal Light.....	67
7.1.2	TACO Sensor, Strong Light.....	69
7.1.3	TACO Sensor, Darkness	70
7.1.4	Kinect Sensor, Normal Light	71
7.1.5	Kinect Sensor, Strong Light.....	73
7.1.6	Kinect Sensor, Darkness	73

List of Figures

Figure 1: Data acquisition time.....	8
Figure 2: The Linearity error (difference of reported to true distance to a reference target, upper part) of the scanner and standard deviation of repeated measurements to the reference target (lower part of the figure) target are often used for performance characterization of TOF systems. Here, blue color denotes the least, red color the most sensitive TOF channel of TACO. The linearization error is below 1 cm and standard deviation between 3 and 8 mm depending on channel sensitivity and target distance, slightly larger than projected and mostly due to laser speckle of the small measurement spot.	9
Figure 3: 3D image of the scanner laboratory. The inset in the lower left corner is a zoom into the lower right part of the full image that demonstrates the possible resolution of the system. The visible line structure demonstrates the choice of mirror trajectories.	10
Figure 4: Standard Deviations & Mean squared errors of tile centres	10
Figure 5: Oxford Technologies Tile test stand as scanned with the TACO sensor using the 1Hz linear mirror plan.....	11
Figure 6: Vertical lines in the range image become jagged. This is not an effect seen when using the point clouds.	11
Figure 7: White wall at 2 meter was measured. a) We see missing points due to low signal received from the wall. b) Intensity data from the same white wall. We observe NO useful information. at Wall_NoLight_Distance_2_Lin_TACODataOutput.nc	11
Figure 8: White flat wall measured at approximately 2 meters distance (measured form the front of the TACO sensor box). A band of data points at 0.05-0.1 meters height is plotted. b) Shadow data is seen in front of the wall, a) the wall is curved. Data used: BareWallRangeTest__200cm_10Hz_20130428T111928TACODataOutput.nc. frame 46.....	12
Figure 9: The four foveation types available for 10 Hz. We see the ringing effect making these mirror plans not usable. By using 1 Hz mirror plans for foveation this ringing does not appear.	13
Figure 10: Mirror plan with error simulation 1 Hz, linear (un-foveated).	13
Figure 11: Mirror plan with error simulation 1 Hz, FOV1 (foveated).	14
Figure 12: Mirror plan with error simulation 1 Hz, FOV2 (foveated).	14
Figure 13: Mirror plan with error simulation 1 Hz, FOV3 (foveated).	14
Figure 14: Mirror plan with error simulation 1 Hz, FOV3 (foveated).	15
Figure 15: Example of 1Hz mirror plan. First column is the linear part of the trajectory and the second column shows the foveated part. Form bottom to top rows Fov1, Fov2, Fov3 and Fov4 trajectories, extracted from a dynamic foveation sequence. These trajectories foveated on the lower, semi-lower, semi-upper and upper parts of the image respectively.	16
Figure 16: Example using 10Hz linear mirror plan.....	17
Figure 17: Type 3 asymmetric trajectories	17
Figure 18: Example using 10Hz foveating mirror plans.....	18
Figure 19: 10 Hz trajectory.	19
Figure 20: 1 Hz trajectory.	19
Figure 21: 10 Hz trajectory.	20
Figure 22: Measurements of a flat wall at approximately 2 meters distance with different illumination properties; room illumination, no external illumination and intense halogen illumination on the scene (the wall).	23
Figure 23: Tiles 1-4 (Top Left – Bottom Right).	25
Figure 24: The Test stand & tiles in the 'Far' position.....	25
Figure 25: Scene & selected points for analysis.	26
Figure 26: Tiles 1-4 (Top-Bottom) scanned with stand in the 'Near' position with 10Hz Linear mirror plan. Each tile is shown in both front-on and side-on views.....	27
Figure 27: Tiles 1-4 (Top-Bottom) scanned with stand in the 'Near' position with 1Hz Linear mirror plan. Each tile is shown in both front-on and side-on views.....	29
Figure 28: Tiles scanned with 1Hz linear mirror plan.....	29
Figure 29: Tile 1 scanned in 'Near' position with Fov3 mirror plan.....	30
Figure 30: Tile with Baseline sample embedded	30
Figure 31: Standard Deviations & Mean squared errors of tile centres.....	30
Figure 32: Clustering results with missing data points.	31
Figure 33: Statistics of analysed points.	31

Figure 34: Unfiltered Data, Averaged data & Median filtered Data of Tile 1 in near position scanned at 1Hz.32

Figure 35: Reference Tile area shown in blue.32

Figure 36: Reduction of Mean squared error through filtering.33

Figure 37: Rear view of centre of Tile 1, scanned with Fov3 1Hz mirror plan & coloured by depth. Due to slight gradient of tile, colour should transition linearly from Red (bottom) – Green (Top).33

Figure 38: Rear view of centre of Tile 1, scanned with 1Hz Linear mirror plan & coloured by depth. Due to slight gradient of tile, colour should transition linearly from Red (bottom) – Green (Top).33

Figure 39: Rear view of centre of Tile 1, scanned with 10Hz Linear mirror plan & coloured by depth. Due to slight gradient of tile, colour should transition linearly from Red (bottom) – Green (Top).34

Figure 40: Sensor data during (top) and after warm-up (bottom).35

Figure 41: Tile scanned face on, rotated in TACO frame X (top & Bottom) & Y (Left & Right).35

Figure 42: Damaged tile surface & artefact.36

Figure 43: Tile Surface detail.36

Figure 44: Covered Tiles36

Figure 45: Covered tiles in 'Near' position scanned with 1 Hz linear mirror plan.37

Figure 46: Clusters detected by pcl Euclidean distance grouping algorithm, with centroids.37

Figure 47: Oscillations along X axis of data points (0s represent no data)38

Figure 48: ABB with vessel like back-drop.39

Figure 49: ABB + end effector with metallic back drop & 10Hz TACO scan of the same scene.40

Figure 50: Black Acrylic tile & TACO image of tile with reflections.41

Figure 51: HiViz material on jacket, TACO scan of jacket face & side on.41

Figure 52: Cotton, Celluloid & Foam rubber spheres & top down TACO scan.41

Figure 53: White plastic foam board target & Taco scan.42

Figure 54: Revised White plastic foam board target.42

Figure 55: Warping of target when viewed side-on.42

Figure 56: Tiles (highlighted in green) & computationally identified positions (highlighted in red).43

Figure 57: ABB scanned with 10Hz & 1Hz mirror plans, note the increased angle in 1Hz scan due to robot motion.43

Figure 58: Public safety test case setup44

Figure 59: Example object cloud.46

Figure 60: Counting true positive points.47

Figure 61: Counting true points.47

Figure 62: 1Hz high density TACO cloud and high density tracked object48

Figure 63: 10Hz low density TACO cloud and high density tracked object49

Figure 64: 3D models used to trained the object detectors.50

Figure 65: TUW setup: Kinect mounted on top of the TACO sensor which is mounted on a mobile platform with a tilting unit. The mobile platform allows us to capture point clouds of the objects at different distances.51

Figure 66: Flat surfaces are reported as wavy surfaces51

Figure 67: 1Hz trajectories do not present wavy planar surfaces but with high laser power, "shadow" effects can be observed (see text for a more detailed explanation).52

Figure 68: Partial views used to train the object detectors. The views are obtained by rendering the 3D model representing the object of interest (i.e. the spray bottle)53

Figure 69: Positive recognition for Kinect and TACO data. Observe the difference in the point cloud reported by both sensors (Kinect: left, TACO: right)54

Figure 70: Correct identification and 6DOF pose estimation for the spray bottle (TACO: top, Kinect: bottom)56

Figure 71: Evolution of the Kinect point cloud under the effect of direct sunlight. The figure shows 8 point clouds (from bottom-left to to top-right) of a static scene captured at different timestamps as the sunlight hits the scene through an opened window. Observe how the dark areas grow as more sunlight hits the field of view of the kinect.57

Figure 72: Point clouds obtained from the taco sensor (left: no sunlight, right: partial sunlight). The image allows seeing the foveated area with a denser resolution (spanning over the objects to be recognized). Observe how the point cloud on the right remains unaltered.57

Figure 73: Point clouds obtained from the Kinect sensor (left: no sunlight, right: partial sunlight).57

Figure 74: The mug and the spray bottle are correctly recognized using the point cloud obtained from the TACO sensor under partial sunlight. The coloured point cloud on the left part of the figure shows the point cloud used for recognition; the rest is discarded. In the right part, the segmented objects are shown with the recognition results overlaid58

Figure 75: Comparison between unfoveated (10Hz linear trajectory) and foveated (Fov 3 - 1Hz trajectory – semi-upper part of the image). Observe the denser band of points on the foveated image. Top: front view, Bottom: side view. In the side view is it again possible to observe the wavy planar surfaces. The effect is mitigated but still visible in 1Hz trajectories.59

List of Tables

Table 1: Standard deviation of distance measurement under various conditions for the ratio P/Pm of laser peak power to maximum laser peak power (approx. 2 kW) and temperature 9

Table 2: Recognition results grouped by distance for Kinect and TACO (1Hz foveated and unfoveated) data53

Table 3: Recognition results grouped by object type.....54

Table 4: Raw recognition results for the different objects used during testing55

1 Executive Summary

This document constitutes the last derivable in WP5 and the TACO project: "Final Test Report". It reports the final status of the actual TACO sensor and software as well as the use case evaluation performed by the end users. Due to the late delivery of the actual hardware, there are some deviations with respect to the description of work, most of them related to:

- Integration and testing of attention/foveation mechanisms using the actual data provided by the TACO sensor.
- Simplification of the use cases according to the availability of the sensor for each end-user (3 weeks).

The document is organized as follows:

- Chapter 2 presents the tests performed on the actual TACO sensor's data (distance measurements, data artifacts, etc.) as well as integration status between the TACO hardware and the foveation software (attention mechanisms and hardware/software communication).
- Chapter 3 reports the findings of the Use Cases deployed at the end-user's sides.
- Chapter 4 evaluates the foveation capabilities of the sensor from the end user's perspective.
- Chapter 5 presents an application map for the TACO sensor and its different sub technologies as well as a summary of the findings gathered during the TACO project.

2 System test and verification

The TACO sensor is characterized in the deliverable D3.3 "First instrument prototype with characterization V1.1" and in TACO deliverable D3.5&D3.4 "Final Prototype Instrument V1.0".

This chapter will present the experiences gained when TACO hardware and TACO foveation software were integrated in April 2013. First we present a short summary of the results achieved. We present information on the range data quality, how several mirror plans enabling foveation are performing, and how attention methods used for performing automatic foveation are working on real TACO data.

2.1 Summary of system integration – hardware/software, system test and verification

The foveation hardware and software was integrated at SINTEF during two weeks in April 2013. The TACO sensor is controlled using predefined mirror plans and performed foveation in real-time based on analysis of the 3D data captured.

Results and observations during the integration work are summarized below:

- **Distance measurements:** Standard deviation of approximately 5mm range.
- **Sensor data artifacts:**
 - Flat walls became curved.
 - Shadow points appear in front of a measured wall at 2 meters distance
 - Stray measurements at the cut-off distance at 0.7 meters
- **External illumination:** The sensor handled direct sunlight fine. Data quality degradation was very limited, both when the scene was exposed to sunlight and when the sunlight shone directly at the sensor. See section 2.6 "Influence of external illumination in the 1550nm band".
- **Mirror plans:**
 - The 10 Hz mirror plans was running, but they have parasitic oscillations which considerably degraded the data quality and the usefulness of the data (see figures in: "Example using 10Hz linear mirror plan" and "Example using 10Hz foveating mirror plans").
 - Foveated mirror plans using the 1Hz trajectories were working without any parasitic oscillations. See section „Example of 1Hz mirror plan“.
 - Dynamic change of mirror plans (i.e. foveation) using predefined plans (one linear and four foveated) was implemented and was working for 1Hz mirror plans see figures in section „Example of 1Hz mirror plan“ (videos are made). And a combination of 10 Hz linear and 1 Hz linear and foveated mirror plans was made. See section "Combination of 10Hz linear and 1Hz foveated trajectories".
- **Attention algorithms:** Several attention methods used for foveation were tested.
- **Realtime performance:** The acquisition time varied between 50-80 ms. Figure below shows the timing plots.

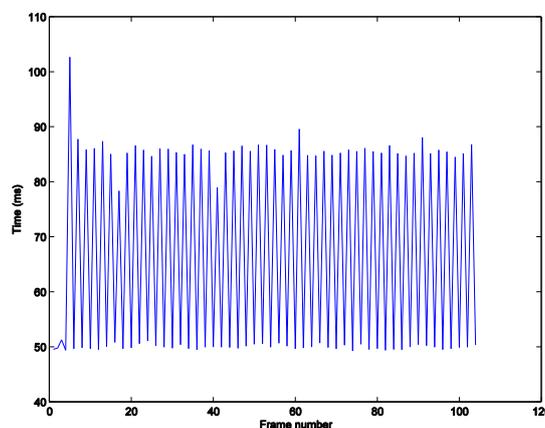


Figure 1: Data acquisition time

2.2 Distance measurements

From D3.5 we found that the standard deviation for point measurements in the time of flight unit is mostly independent of temperature and other parameters as, e.g., laser power or target properties. To obtain values spanning the entire dynamical range we employ a grey wedge attenuator in the light path. Table 1 shows representative measurement results of the standard deviation in the four channels. The contributions of the two TDCs vary slightly in the range 0.1 to 1 mm. Apart from the results for the most sensitive channel 0, the results coincide with our expectations from the design phase (see D2.1).

Parameter	channel 0 most sensitive	channel 1	channel 2	channel 3 least sensitive
P/P _m =1/4, 39 °C (opt. attenuation)	6.0 mm	4.1 mm	4.2 mm	3.7 mm
P/P _m =1/4, 30 °C (opt. attenuation)	6.1 mm	4.7 mm	4.5 mm	4.1 mm
P/P _m =1/2, 39 °C	4.6 mm	4.7 mm		
P/P _m =1, 39 °C				3.4 mm

Table 1: Standard deviation of distance measurement under various conditions for the ratio P/P_m of laser peak power to maximum laser peak power (approx. 2 kW) and temperature

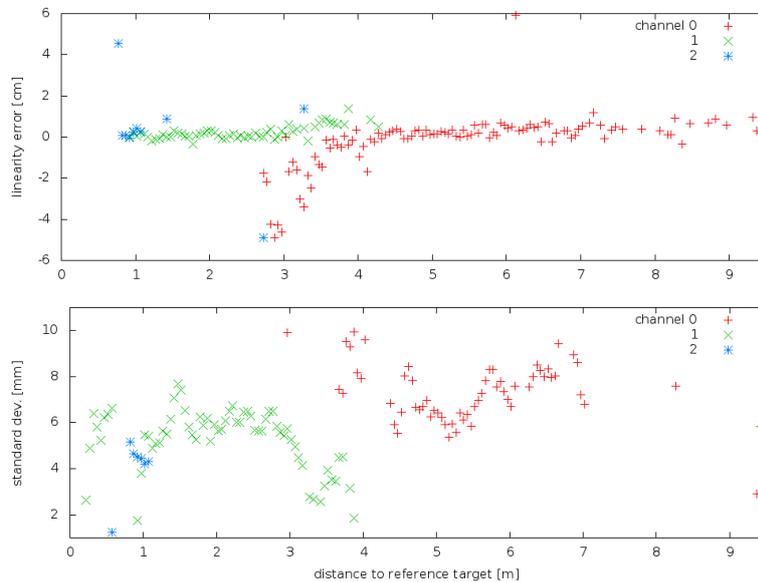


Figure 2: The Linearity error (difference of reported to true distance to a reference target, upper part) of the scanner and standard deviation of repeated measurements to the reference target (lower part of the figure) target are often used for performance characterization of TOF systems. Here, blue color denotes the least, red color the most sensitive TOF channel of TACO. The linearization error is below 1 cm and standard deviation between 3 and 8 mm depending on channel sensitivity and target distance, slightly larger than projected and mostly due to laser speckle of the small measurement spot.

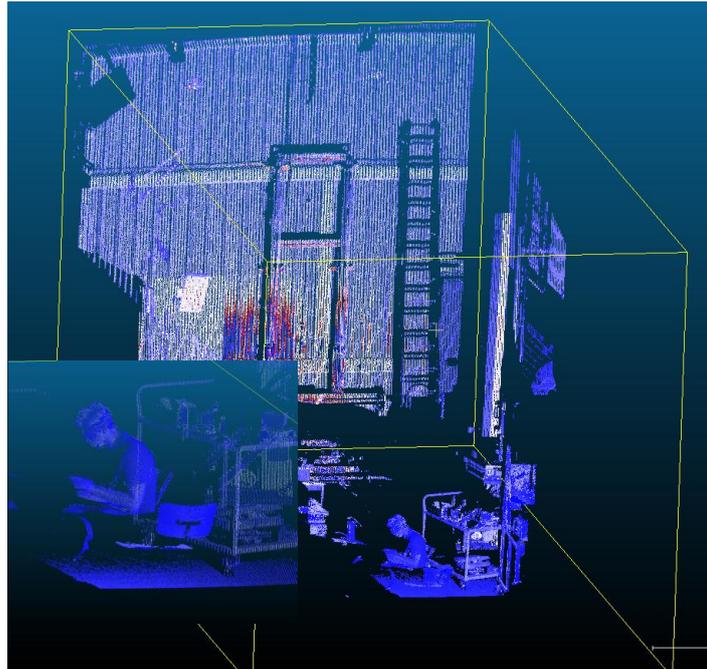


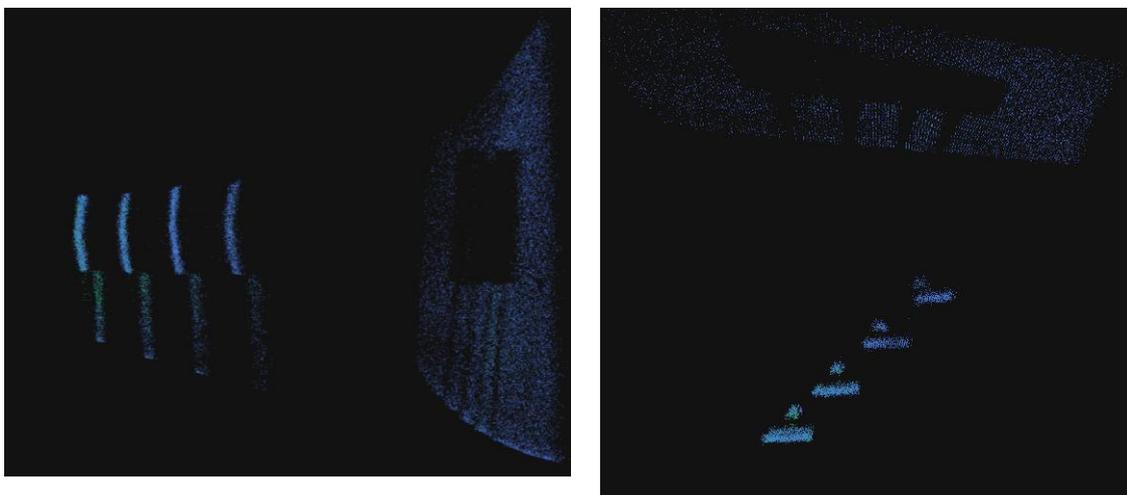
Figure 3: 3D image of the scanner laboratory. The inset in the lower left corner is a zoom into the lower right part of the full image that demonstrates the possible resolution of the system. The visible line structure demonstrates the choice of mirror trajectories.

A comparison of the measured & actual positions of the centres of the Tiles scanned with TACO at Oxford Technologies (where available, see 3.1.1.4.1) was conducted and it was found that the sensor returned the following statistics:

	1Hz Linear	10Hz Linear
Z axis Standard Deviation	0.012917157	0.010923139
Z axis Mean Squared Error	0.000171236	0.000135656

Figure 4: Standard Deviations & Mean squared errors of tile centres

For comparison with Figure 3 see the actual scans of the Oxford Technologies Tile test stand:



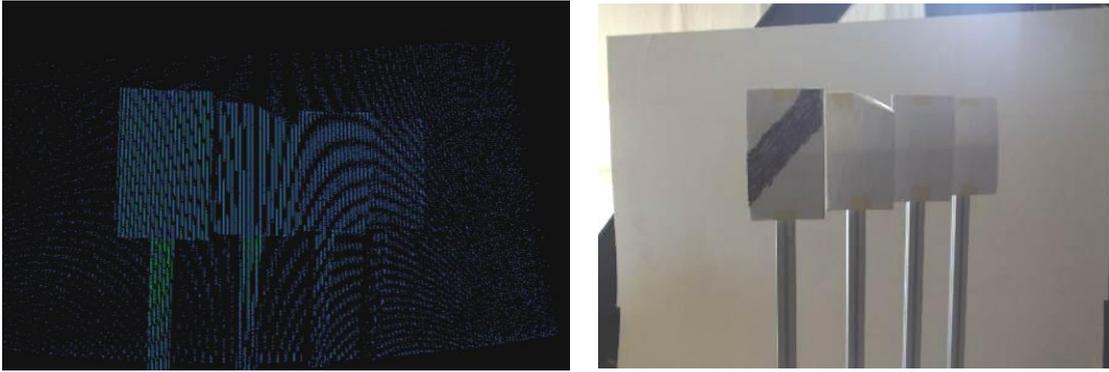


Figure 5: Oxford Technologies Tile test stand as scanned with the TACO sensor using the 1Hz linear mirror plan

2.3 Observed Sensor Data Artifacts

In this section we give an overview of the artifacts observed when using the sensor and when doing foveation.

- In the range images straight vertical lines became jagged. And the amount of jaggedness was varying from frame to frame. The point clouds however were fine. See Figure 6.

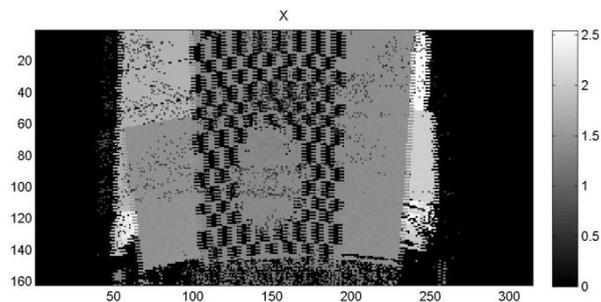


Figure 6: Vertical lines in the range image become jagged. This is not an effect seen when using the point clouds.

- There was a fair amount of missing data due to low signal reflected by the scene. (xyzi data are then set to 0, zero). See Figure 7.
- The intensity information from the sensor was very noisy and was therefore not suitable for analysis (see Figure 7).

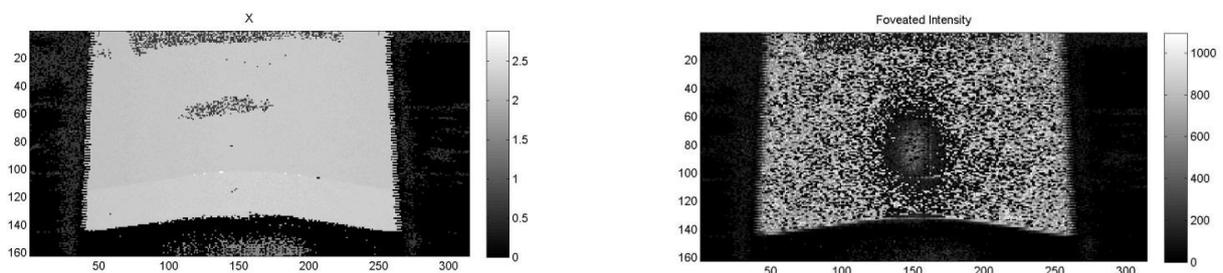


Figure 7: White wall at 2 meter was measured. a) We see missing points due to low signal received from the wall. b) Intensity data from the same white wall. We observe NO useful information. at Wall_NoLight_Distance_2_Lin_TACODataOutput.nc

- A flat wall at 2 meters distance had a slight curvature in the point cloud data. See Figure 8a).
- At short distances (2m), we observed a "shadow" of points some centimeters (approx. 5cm) in front of the actual object measured. See Figure 8b). At longer distances, this did not seem to occur.

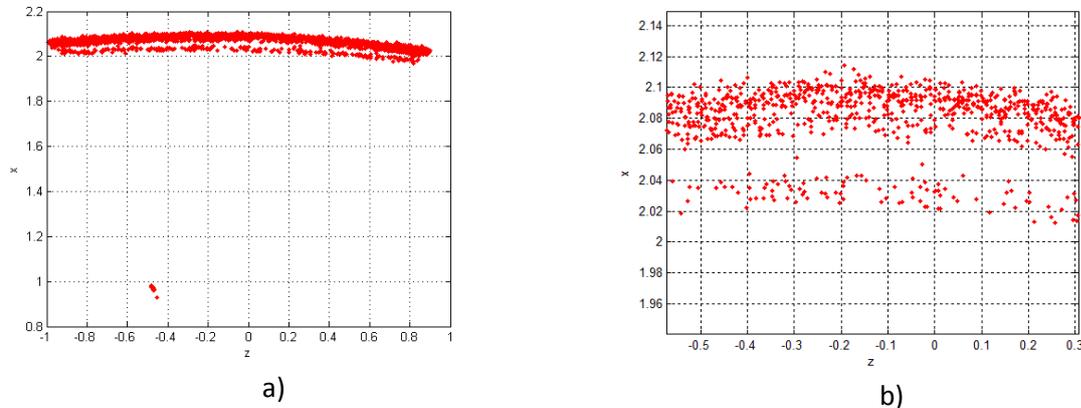


Figure 8: White flat wall measured at approximately 2 meters distance (measured from the front of the TACO sensor box). A band of data points at 0.05-0.1 meters height is plotted. b) Shadow data is seen in front of the wall, a) the wall is curved. Data used:

BareWallRangeTest__200cm_10Hz_20130428T111928TACODataOutput.nc. frame 46

- When we received the sensor, we observed that a significant portion of the measurements were erroneously reported at 0.7 meters. This was an artifact and was removed by improving the module of the sensor software. When the sensor was shipped from SINTEF this was no longer an issue. The reported distance measurements are dependent on the operating temperature of the sensor. The most accurate measurements were observed when sensor had reached operating temperature.

2.4 Changing mirror plans

The TACO sensor can capture data using several predefined mirror plans. The available mirror plans are:

- One 10Hz linear mirror plan
- One 1Hz linear mirror plan
- Four 1Hz foveating mirror plans

See Figure 9 to Figure 14 for plots of the mirror plans.

The reasons for having predefined mirror plans was given in deliverable D3.3, and is due to the mirror motion as reported by the piezoresistive circuitry cannot be used directly to reconstruct images because of the large amount of electronic noise of uncertain origin in the values that the embedded controller obtains from the AD converters on the mirror controller board. The effect is that on repetition of the scanning trajectory the angles associated with the same target point vary quite substantially, leading to an uncertainty in 3D space substantially larger than the uncertainty in the distance measurement.

10Hz foveation mirror plans were tested and these had severe ringing. Due to this we used 1Hz or 10Hz un-foveated mirror plans and only 1Hz foveated mirror plans. We elaborate more on this in the following sections.

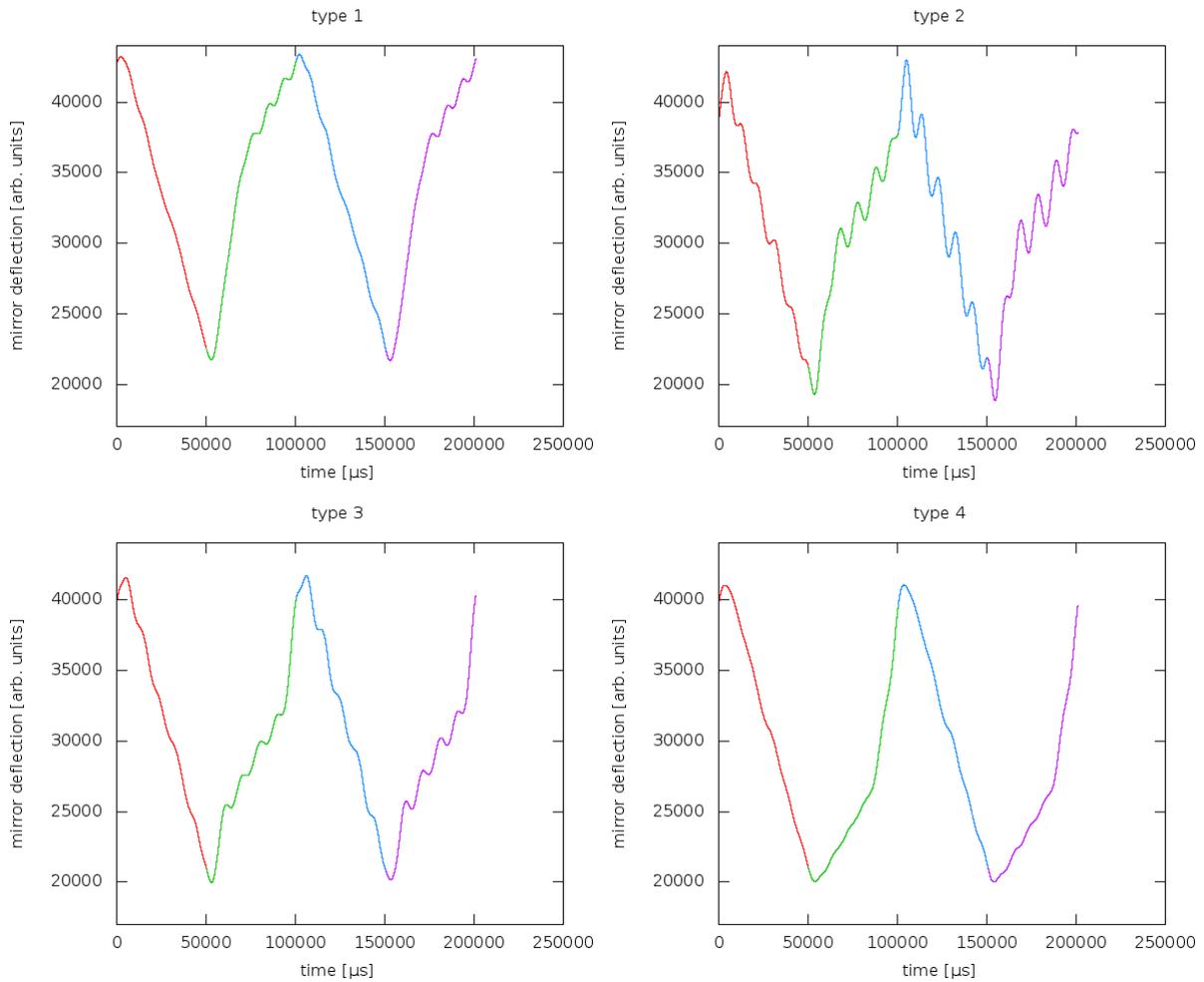


Figure 9: The four foveation types available for 10 Hz. We see the ringing effect making these mirror plans not usable. By using 1 Hz mirror plans for foveation this ringing does not appear.

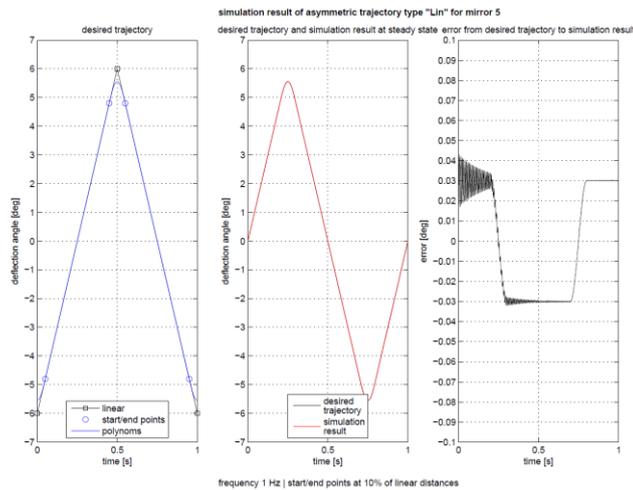


Figure 10: Mirror plan with error simulation 1 Hz, linear (un-foveated).

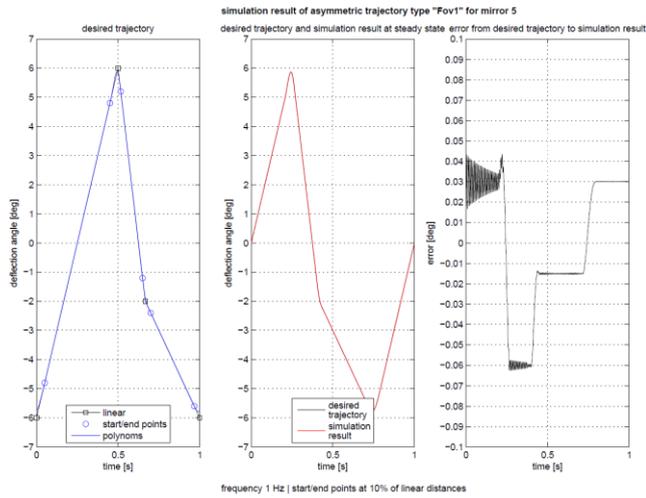


Figure 11: Mirror plan with error simulation 1 Hz, FOV1 (foveated).

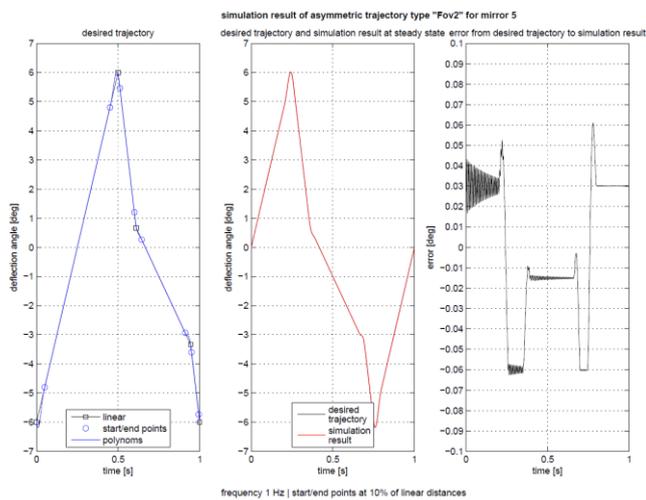


Figure 12: Mirror plan with error simulation 1 Hz, FOV2 (foveated).

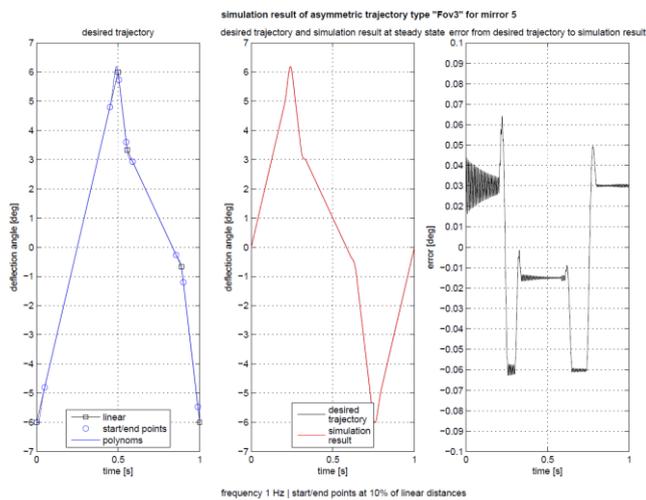


Figure 13: Mirror plan with error simulation 1 Hz, FOV3 (foveated).

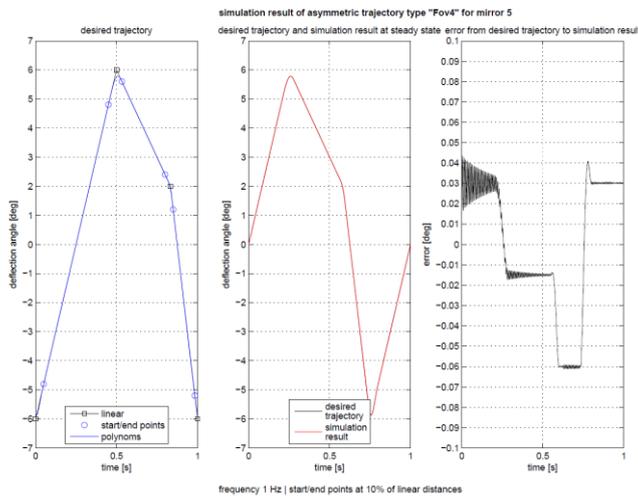
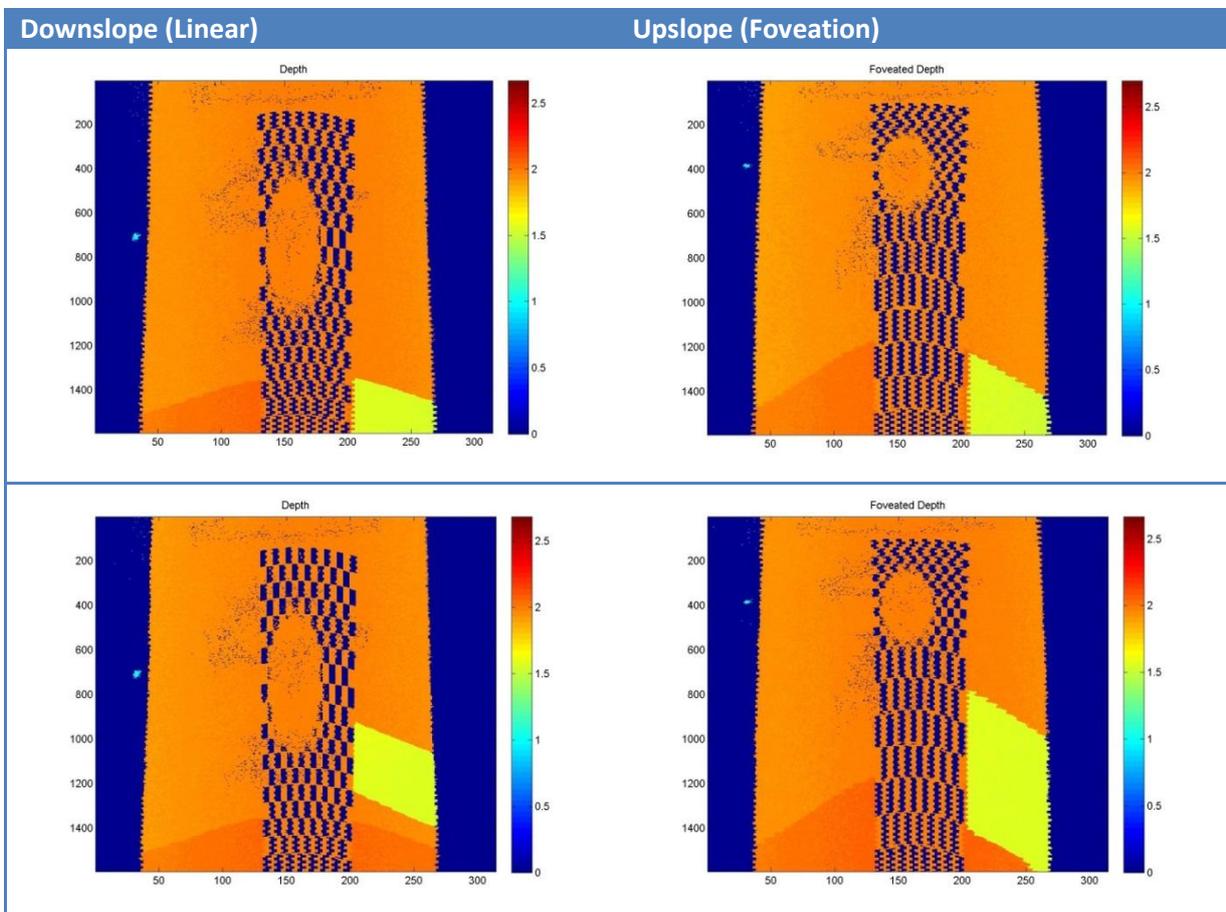


Figure 14: Mirror plan with error simulation 1 Hz, FOV3 (foveated).

2.4.1 Example of 1Hz mirror plan

The 1Hz foveation does not exhibit parasitic oscillations as seen from the figures below. In this figure we show the results of the Fov1, Fov2, Fov3 and Fov4 trajectories, extracted from a dynamic foveation sequence. These trajectories foveated on the lower, semi-lower, semi-upper and upper parts of the image respectively.



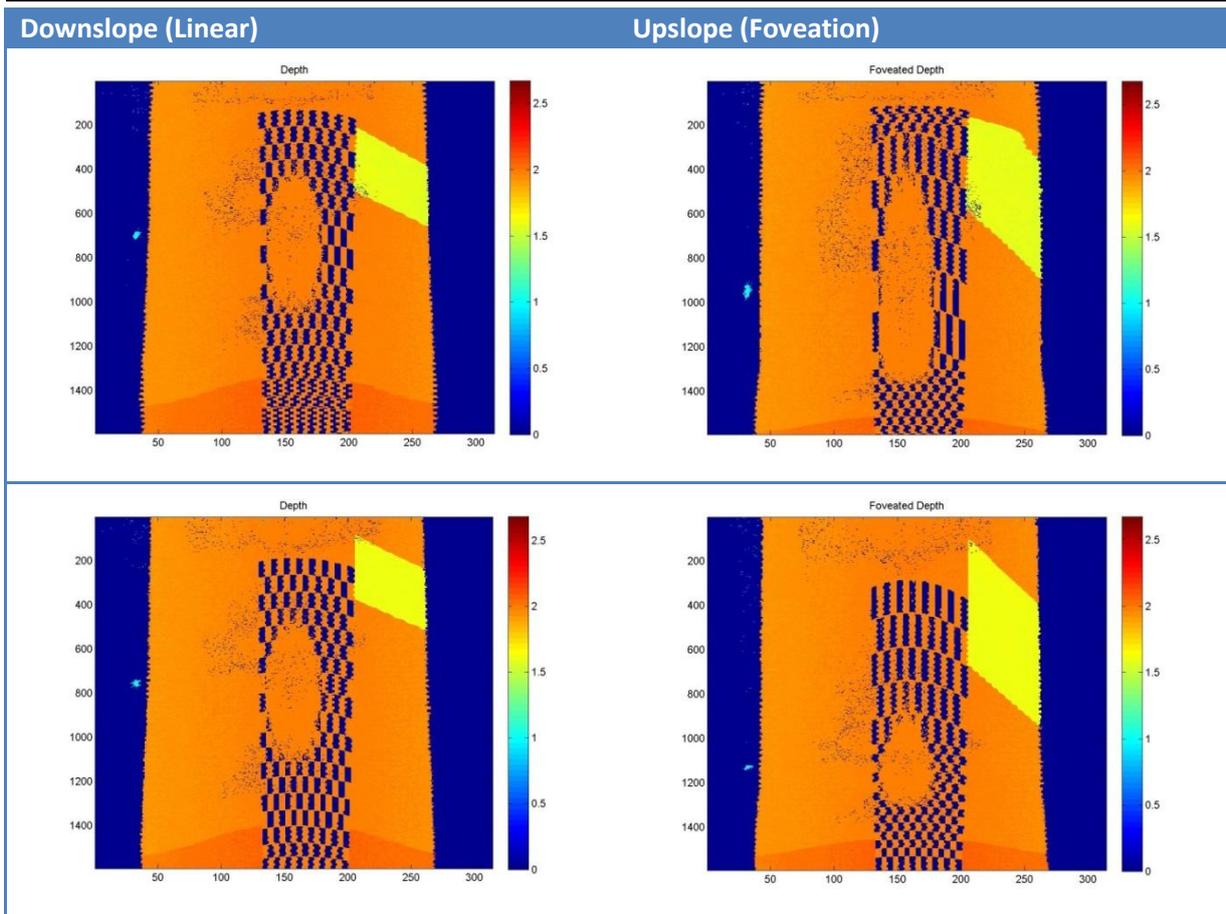
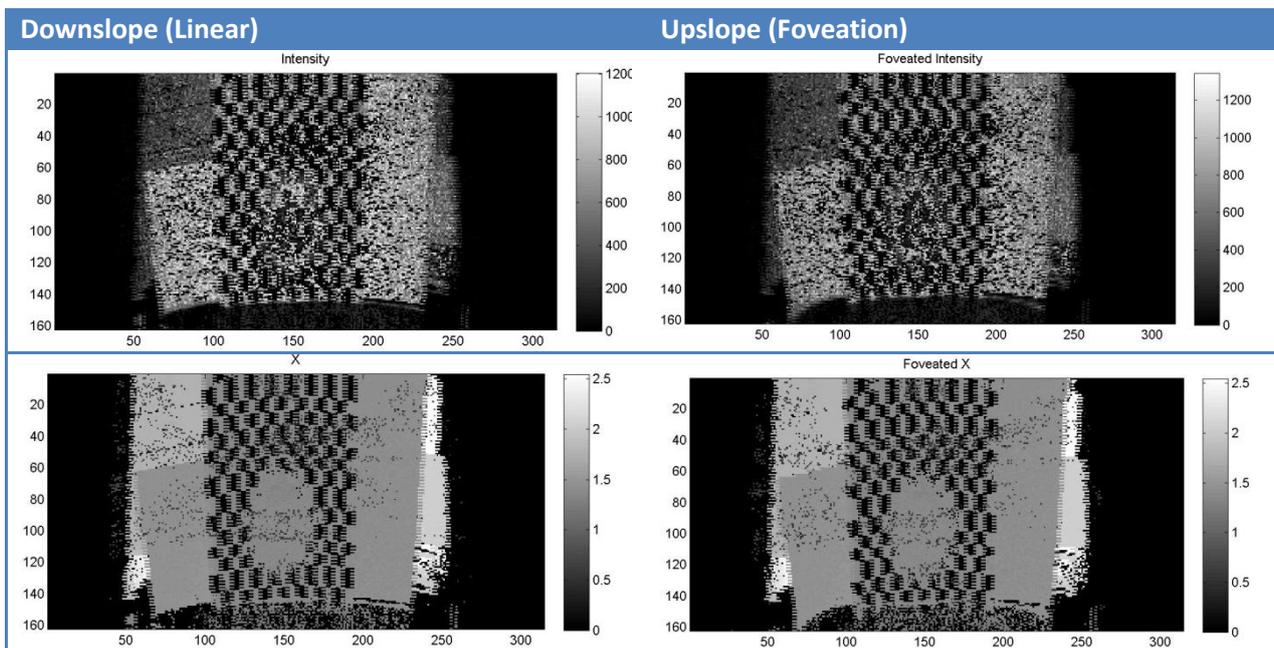


Figure 15: Example of 1Hz mirror plan. First column is the linear part of the trajectory and the second column shows the foveated part. From bottom to top rows Fov1, Fov2, Fov3 and Fov4 trajectories, extracted from a dynamic foveation sequence. These trajectories foveated on the lower, semi-lower, semi-upper and upper parts of the image respectively.

2.4.2 Example using 10Hz linear mirror plan

Seemed to work well – we did not observe any oscillations. See the figures below.



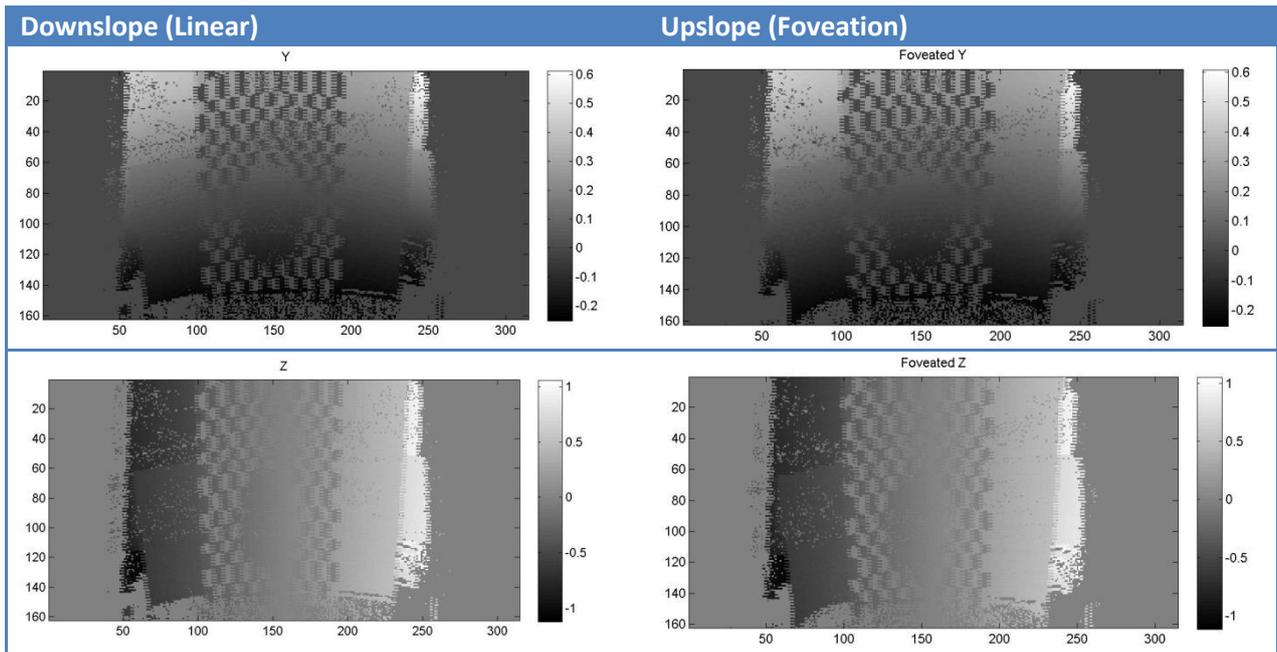


Figure 16: Example using 10Hz linear mirror plan

2.4.3 Example using 10Hz foveating mirror plans

For this mirror plan we observed strong parasitic oscillations in both linear and foveated images using this mirror plan. We see these as a "wavy" change in resolution across the foveated image and as oscillations in the trajectory plot. See the figures below.

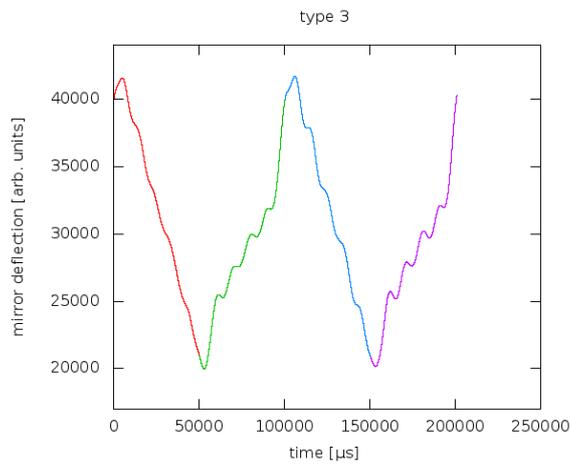


Figure 17: Type 3 asymmetric trajectories

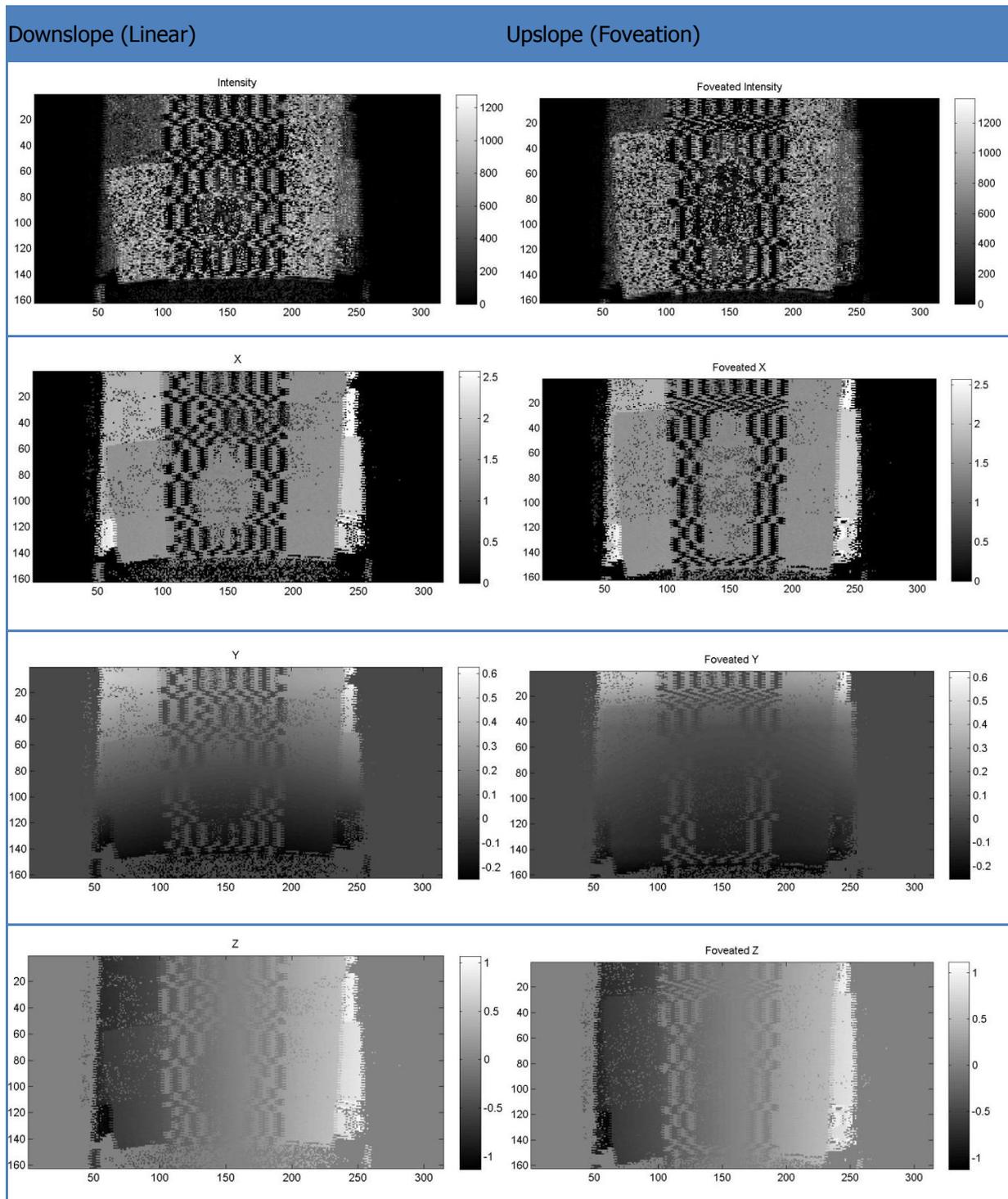


Figure 18: Example using 10Hz foveating mirror plans

2.4.4 Combination of 10Hz linear and 1Hz foveated trajectories

To resolve the issue of parasitic oscillations and to make foveation really boost the data sampling density, we implemented foveation as a combination of 10 and 1Hz trajectories – 10 Hz unfoveated data in combination with 1 Hz foveated data.

The users can also give as input to the system which of the predefined mirror plans to use. Then NO automatic foveation will be used.

Below are snapshots from a video made by Shadow in the beginning is in 10Hz linear mode, then changes to 1Hz_Fov1 mode (and then we do the object detection), then changes back to 10Hz linear (and we pick the object and do the tracking).



Figure 19: 10 Hz trajectory.



Figure 20: 1 Hz trajectory.



Figure 21: 10 Hz trajectory.

2.5 Attention algorithms used for foveation

To select where to foveate several attention features were developed in work package 4 of the TACO project. Based on the TACO use cases, we made a prioritization on which features to test on real TACO sensor data.

We have tested the following dynamic foveation algorithms on data from the TACO sensor (online while interfacing with the sensor and offline on saved data files).

2.5.1 Range model ('attentionConfig_RangeModel')

Describing the feature: Objects moving in the range image can be detected by observing a deviation between a new image and a reference image. The main idea of the range model feature is to estimate the expected range and noise in the scene based on historical data. This feature approximates the average of each range pixel and corresponding tolerance measure as a moving average with exponential decay. The benefit of modeling range images is that pixel changes relate directly to motion in the scene, both due to camera and object movement. By modeling the expected range in each pixel (termed the background), objects moving in the scene can be detected at pixel level as a deviance from this background model.

The method:

- is running in realtime
- works to some extent on the data and parameters needs tuning

Conclusion: is not working due to no time available for parameter tuning (and not sufficient interest from end users)

2.5.2 Contour tracker ('attentionConfig_Tracking')

Describing the feature: The tracker will track foreground objects that are rather well separated in depth from the background. The tracker uses depth data alone for tracking.

The method:

- is running in realtime

- does the initial segmentation of objects fine
- loses track of the object rapidly. This is assumed to be due to
 - large amount of missing data from sensor
 - jagged edges in the sensor data
 - might also be due to the noise in the range data

Conclusion: this method does not work sufficiently well to be of sufficiently practical use.

2.5.3 Distance based foveation ('attentionConfig_Distance')

Describing the feature: segmenting all data points based in a given distance from sensor, between a maximum and a minimum distance.

The method:

- is running in real-time
- used to demonstrate dynamically changing mirror plans

Conclusion: a simple method that works well enough to showcase dynamic foveation

2.5.4 Other foveation methods

Foveation algorithms that have not been tested (due to time pressure) at SINTEF, but that we still have an opinion on how will perform on the TACO sensor:

- lineMod (6DOF pose information of tracked CAD object):
 - not tested
 - **Conclusion:** will probably not work due to poor quality of intensity data
- Tracking of retro reflective markers (6DOF pose information of marked object):
 - the markers is not seen in the intensity image due to poor quality of intensity data + difficult to find good retroreflectors at 1550 nm
 - **Conclusion:** not possible to use this method due to poor quality of intensity data
- TUW's bottom up foveation (available as open source Toolbox):
 - not tested on real data
 - algorithm was tested on simulated data (see D4.3 and D4.4)
 - **Conclusion:** will probably not work due to missing points on the real data coming from the sensor
- TUW's object based foveation:
 - not tested on real data
 - algorithm was tested on simulated data (see D4.3 and D4.4), no changes since then
 - **Conclusion:** more time would be needed to test the algorithm on real data and see if any adaptations are needed.

2.5.5 Sending mirror plans over the network from robot to the TACO sensor

The robot or the end user can also decide which mirror plans to use by using the text based protocol to send commands to the TACO sensor (protocol is defined in D4.3). Then there will be no foveation based on foveation algorithms and the mirror plan used is the same until a new mirror plan is sent to the sensor.

Command used: mirrorPlanName <timestamp> <mirror plan name>

2.6 Influence of external illumination in the 1550nm band

In order to check how robust the TACO sensor's measurements are we tested the sensor under influence of external light sources. The laser in the TACO sensor's time of flight unit has wavelength 1550nm, and the tests were performed using external illumination consisting of light with this wavelength.

From these tests we can conclude that the TACO sensor is not influenced by external light sources with the strengths tested here.

Acquisition setup and tests:

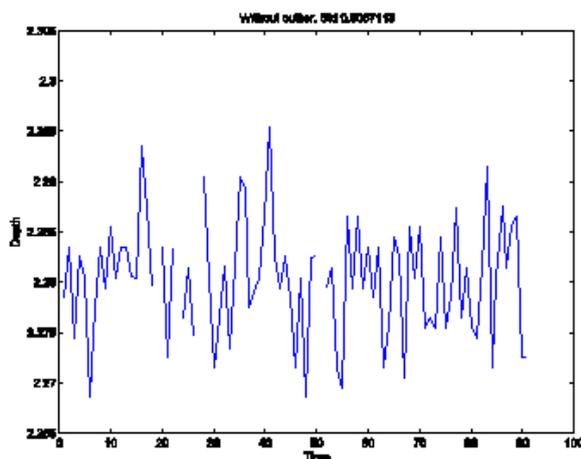
- The TACO sensor was placed at a distance of 2 meters from a flat surface (a wall).
- We acquired 100 frames of data, both up slope and down slope is included (jaggedness influence the data)
- Temporal evaluation of data/noise was performed

Results:

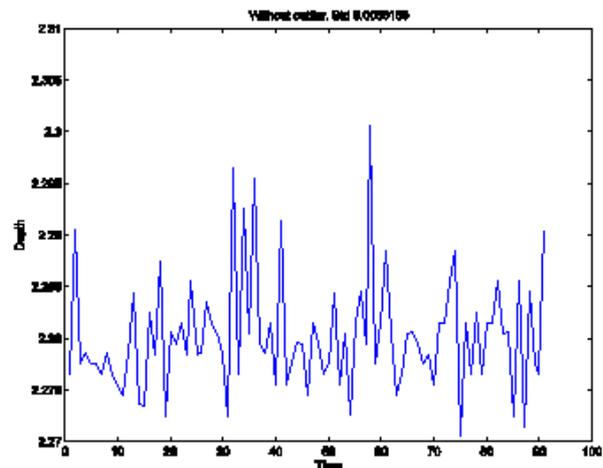
Outliers due to the sensors cut-off at 0.7 meters and the shadow points in front of the wall caused a large standard deviation before outlier removal (in the cm range).

- In regular room illumination:
 - Over a 9x9 neighborhood and 100 frames: Depth standard deviation 5.6 cm.
 - At one pixel without outliers: Depth standard deviation of 5.7 mm.
- Without illumination:
 - Over a 9x9 neighborhood and 100 frames: Depth standard deviation 6.4 cm.
 - At a pixel without outlier: Depth standard deviation of 5.5 mm.
- Illuminated by halogen light (100W):
 - Over a 9x9 neighborhood and 100 frames: Depth standard deviation 3.4 cm.
 - At one pixel without outliers: Depth standard deviation of 6.7mm.

Room Illumination



No External Illumination



Halogen Illumination

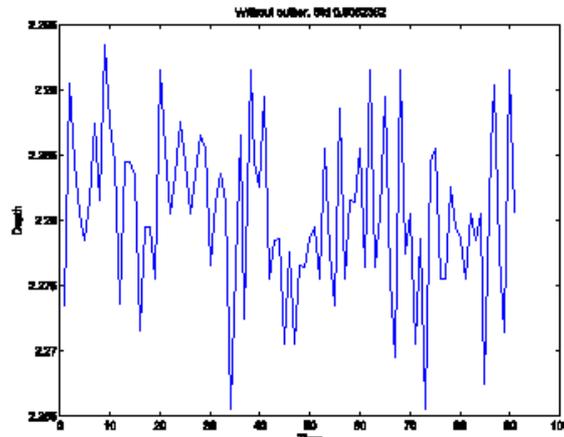


Figure 22: Measurements of a flat wall at approximately 2 meters distance with different illumination properties; room illumination, no external illumination and intense halogen illumination on the scene (the wall).

3 Use Cases Evaluation

3.1 3D Inspection & Augmented Reality (OTL)

The first use case to be performed by OTL was the observation of typical plasma & mechanical damage to a set of tiles emulating the typical tiles found within a fusion reactor. The aim of this test was to observe the damage with a high scan rate & then to foveate on the damaged regions, demonstrating the TACO's ability to save time during operations by providing higher scan resolutions and averaging on areas of interest.

The second use case was to observe targets representing plasma vessel components (in both unfoveated modes) as they are manipulated by a robotic device & to track these objects within the scene. The aim of this test is to prove that the sensor can be used for such tracking activities that would then be used to create an augmented reality display for an operator of the robotic device.

3.1.1 3D Inspection

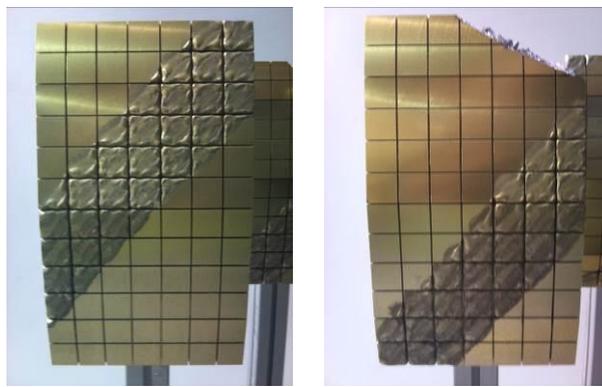
3.1.1.1 Set-up

The sensor was placed at a known position relative to a test stand upon which 4 tiles with various types of damage were present. The range and intensity data of the scene were recorded in all available foveation modes. This process was repeated to collect data across a large region of the sensor's range. This data was then used to estimate the tile positions and observe the damage in either the intensity or positional data. In addition, post processing was carried out to attempt to improve the quality of the data.

3.1.1.2 Technical description of the method

The targets for this test were 4 tiles emulating those found within a plasma vessel environment. The tiles were constructed of 6082T aluminium, anodized to provide a surface similar to the beryllium tiles typically used in such an environment, two tones of gold anodizing were applied to emulate the differing finishes on tiles. The tiles were manufactured with a castellated surface, typical of fusion reactor tiles. The tiles were burned with a plasma torch to emulate the surface damage due to fusion plasma discharge. In addition a corner was removed from one tile to emulate extreme plasma/mechanical damage to a tile.

The tiles were placed on a test stand at varying distances from a plane parallel to the faces of the tiles to permit the imaging of tiles at multiple ranges from the sensor in a single test run. The Test stand was placed at ranges: 1.1m (Near), 1.71m (Mid) & 2.29m (Far) from the Sensor.



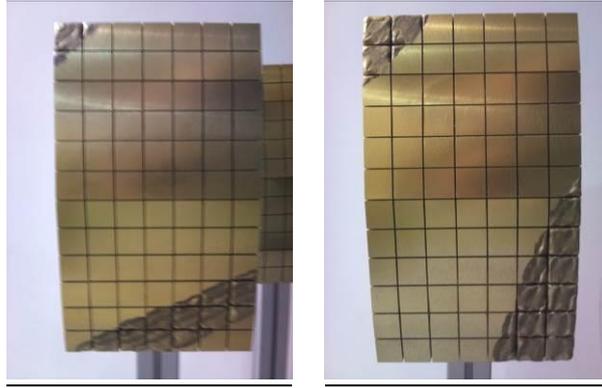


Figure 23: Tiles 1-4 (Top Left – Bottom Right).

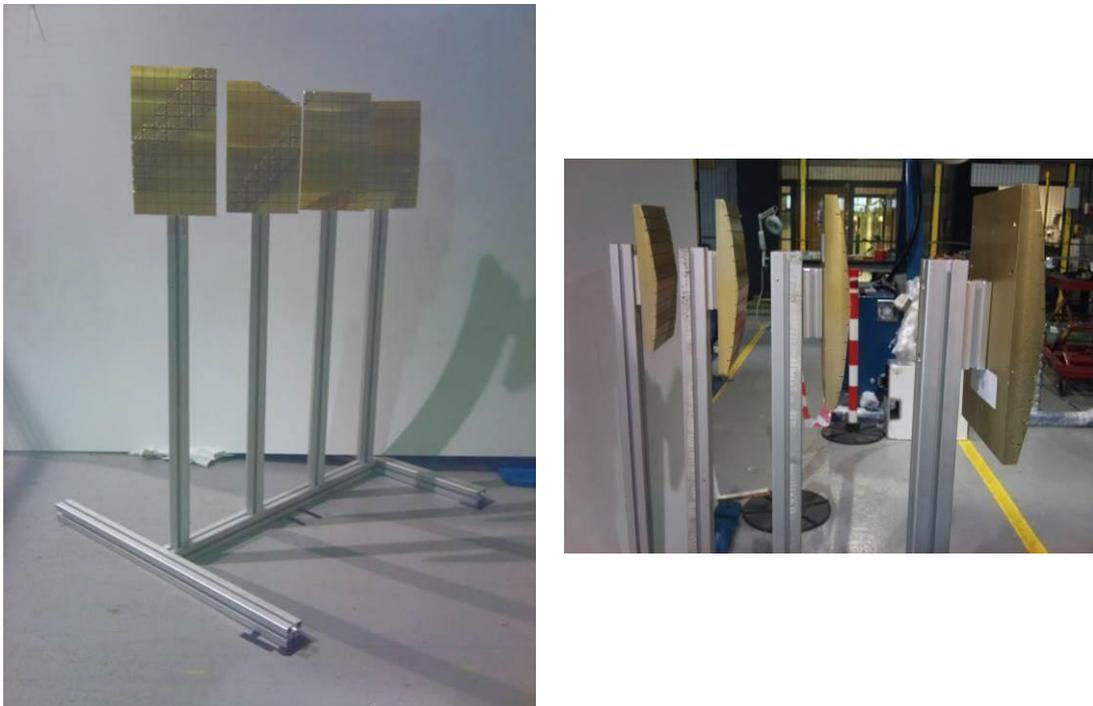


Figure 24: The Test stand & tiles in the 'Far' position.

The Data was collected utilizing the TACO ROS node & post processed utilizing the Willow Garage point cloud library (pcl).

As well as using the 'raw' sensor data, in order to reduce the noise observed within a static scene an average of the observed data over a 10 second duration was taken for each data point.

3.1.1.3 Evaluation method

3.1.1.3.1 Feature Detection

The Range & intensity data was inspected visually to confirm that tile damage could be observed by an operator or robotic control system viewing a scene with the TACO sensor. This was performed in all applicable foveation modes to show any improvement in image quality produced by applying the foveation methods.

The data collected in each foveation mode was compared with the measured centre of the tile, additionally the data was filtered & clustered to extract the computationally measured centre point & dimensions of the tiles for comparison with the actual. To assess any sensor noise several points were selected on the tile surface as well as the wall in the background & the variations in the data analysed for approx. 10s of scans.

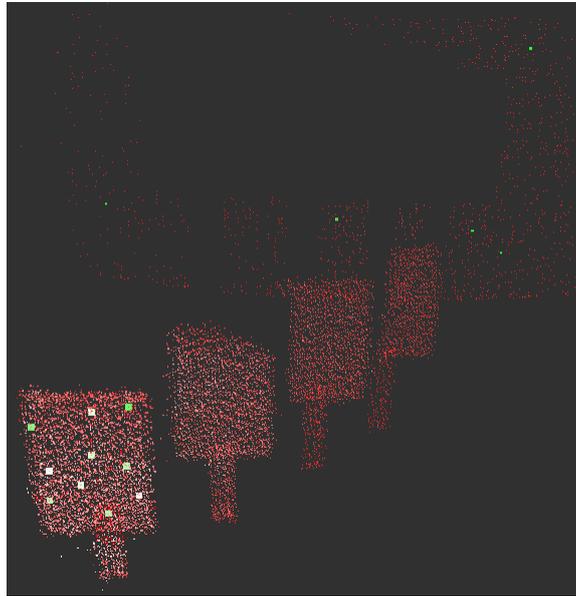


Figure 25: Scene & selected points for analysis.

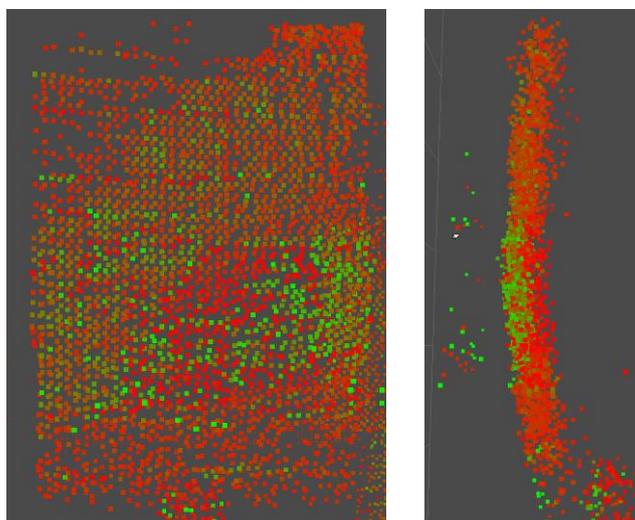
3.1.1.3.2 Averaging of point cloud data

The point cloud position & intensity data was averaged over several scans & filtered in order to observe the effect of minimizing noise present in the data. Visual checks & the change in mean squared error & std. deviation in the sensor's Z dimension were used to provide a measure of the improvement. To make the output metrics clearer, the reference surface was located centrally on the tile's measured position.

3.1.1.4 Results

3.1.1.4.1 Feature Detection

The data observed by scanning the tiles is presented below:



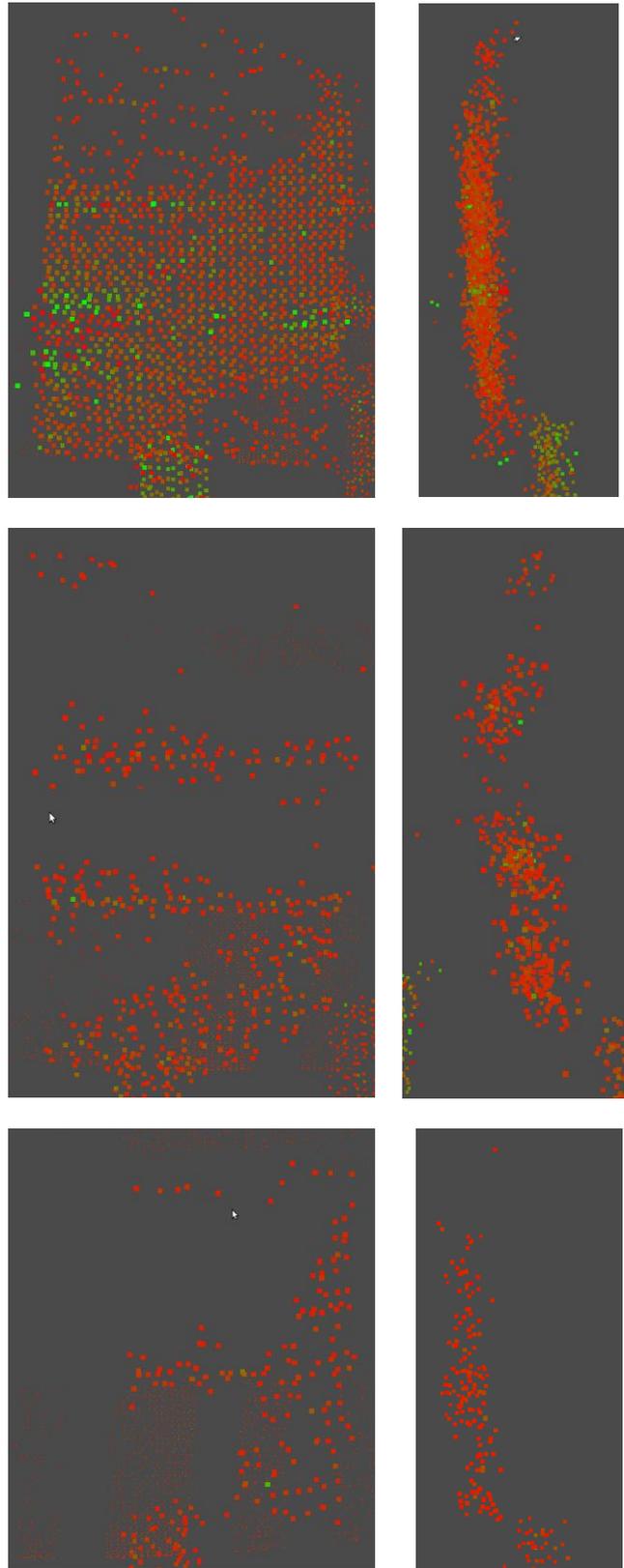
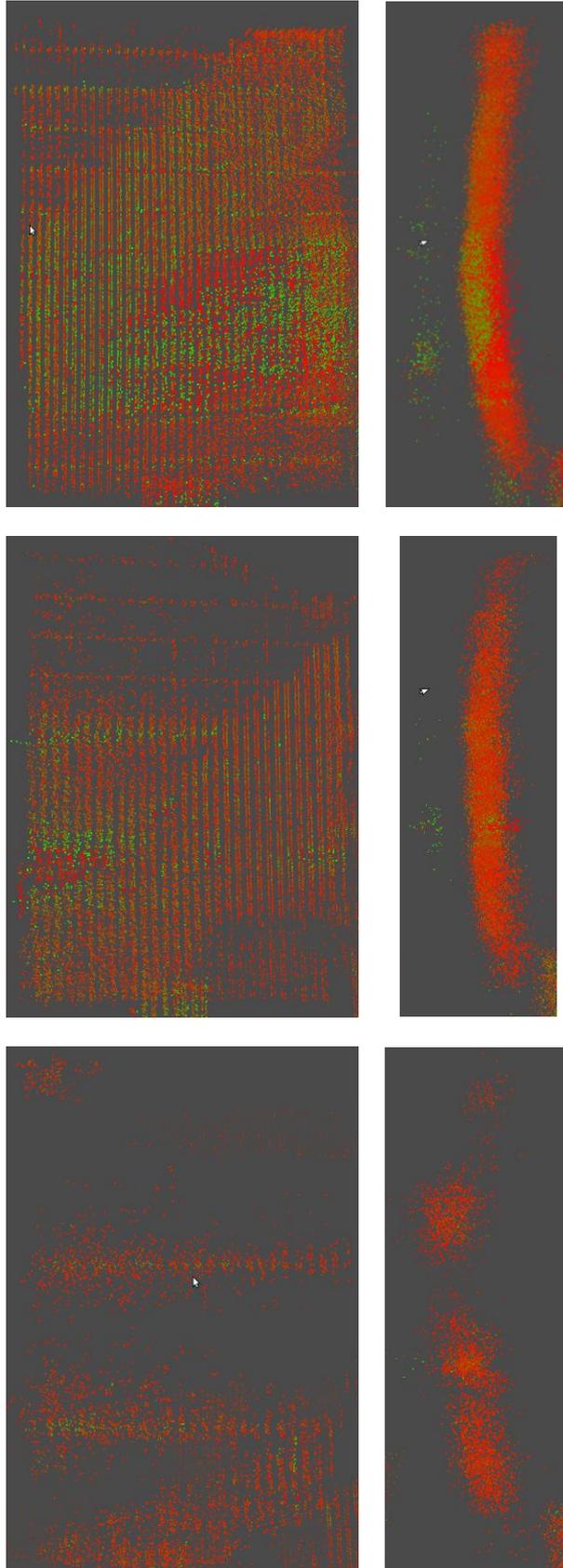


Figure 26: Tiles 1-4 (Top-Bottom) scanned with stand in the 'Near' position with 10Hz Linear mirror plan. Each tile is shown in both front-on and side-on views.



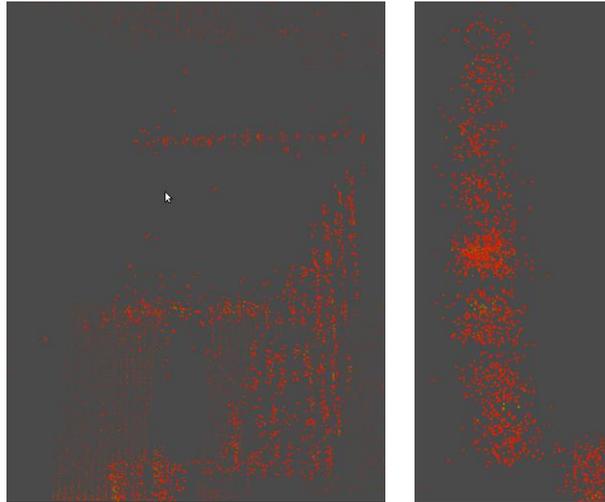


Figure 27: Tiles 1-4 (Top-Bottom) scanned with stand in the 'Near' position with 1Hz Linear mirror plan. Each tile is shown in both front-on and side-on views.

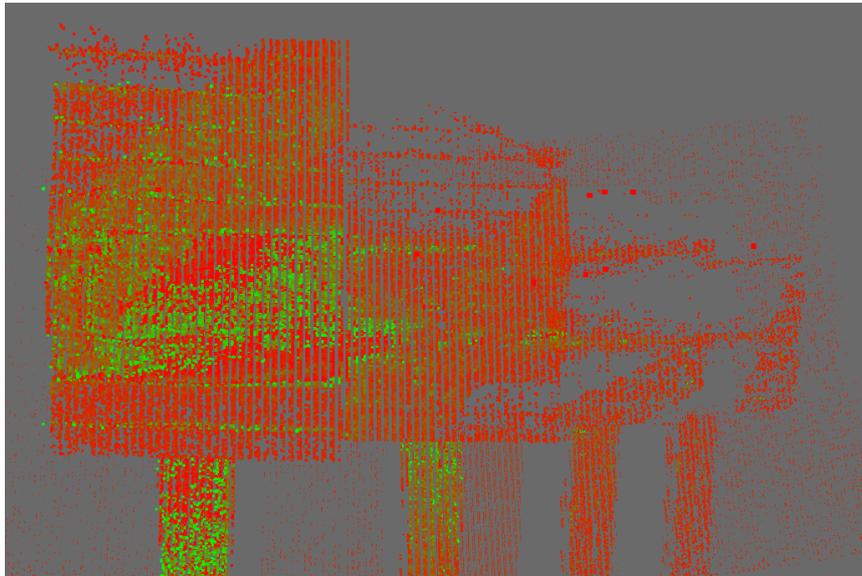


Figure 28: Tiles scanned with 1Hz linear mirror plan.

The sensor data exhibited several issues that would prevent its use in RH operations:

- Missing data points over large areas of the tile surface.
- Small amounts of noise in front of tiles.
- Lack of correlation between damaged tile surfaces & intensity data.
- High Noise in Z dimension data.

The foveation did work as expected, delivering higher density data points within the scene when the mirror plan is changed. This data was still subject to the same disturbances as the unfoveated data.

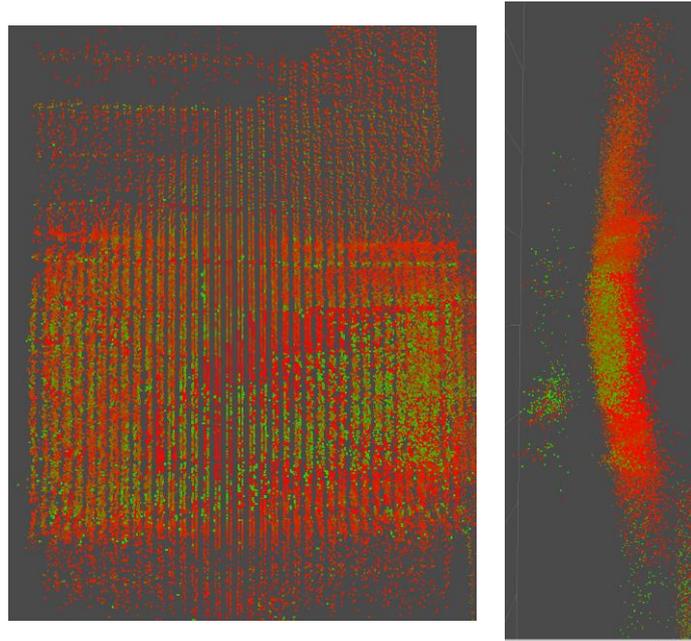


Figure 29: Tile 1 scanned in 'Near' position with Fov3 mirror plan.

Due to the presence of missing data on the tiles' surfaces, only Tile 1 could be analysed effectively. A sample from the centre of the tile was taken & compared to the known baseline.

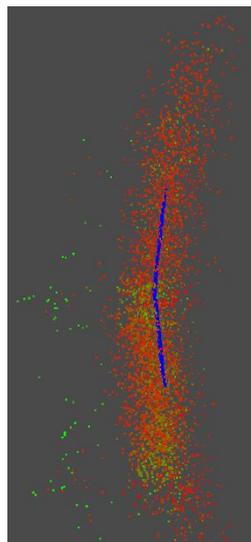


Figure 30: Tile with Baseline sample embedded

Analysing the sensor in this way with all viable data collected in the applicable foveation modes yielded these results:

	1Hz Linear	10Hz Linear
Z axis Standard Deviation	0.012917157	0.010923139
Z axis Mean Squared Error	0.000171236	0.000135656

Figure 31: Standard Deviations & Mean squared errors of tile centres

Due to the missing data it proved challenging to computationally locate the centre of the tiles as the centroids of the data collected were offset. See the below for a typical output of the clustering algorithm:

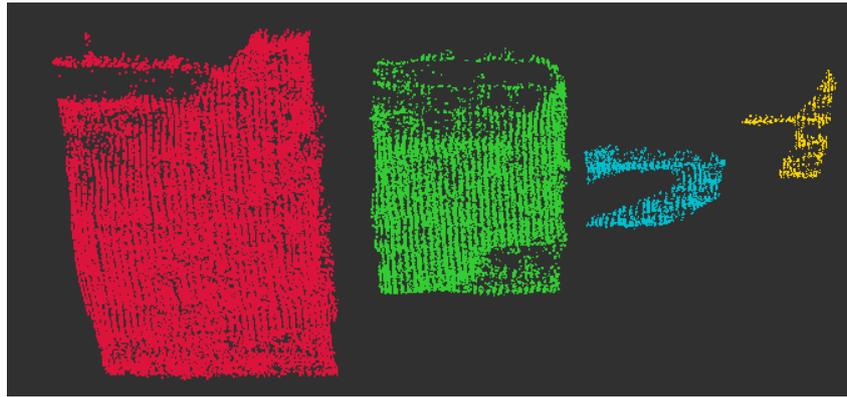


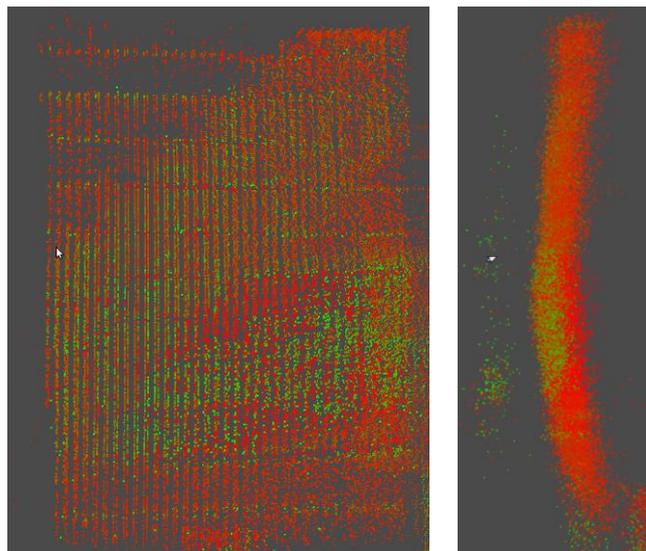
Figure 32: Clustering results with missing data points.

The amount of noise that we observed across the data points (sampled with the 10Hz mirror plan) is categorized in the table below.

10 points on tile				
Axis	X	Y	Z	
Average variance (in m ²)	3.89e-6	4.19e-7	7.2e-5	
Average standard deviation (in m)	0.001802	0.00054	0.00759	
Average standard error (in m)	1.34e-5	4.22e-5	5.8e-5	
10 points on wall				
Axis	X	Y	Z	
Average variance (in m ²)	0.000022	0.000001	0.000097	
Average standard deviation (in m)	0.004616	0.000987	0.009745	
Average standard error (in m)	0.000106	0.000023	0.000240	

Figure 33: Statistics of analysed points.

3.1.1.4.2 Averaging of point cloud data



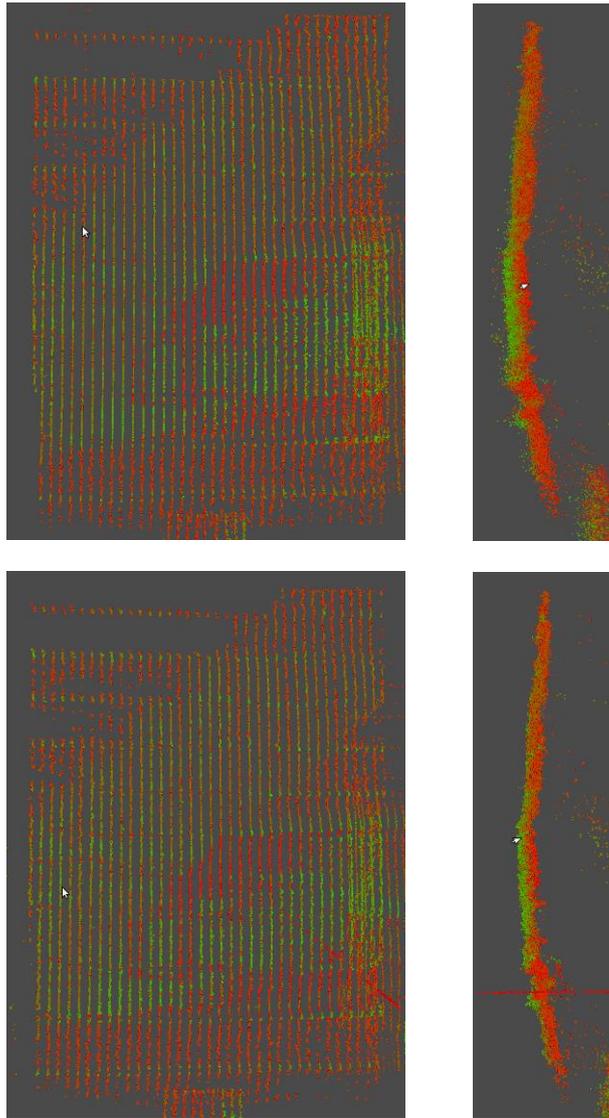


Figure 34: Unfiltered Data, Averaged data & Median filtered Data of Tile 1 in near position scanned at 1Hz.

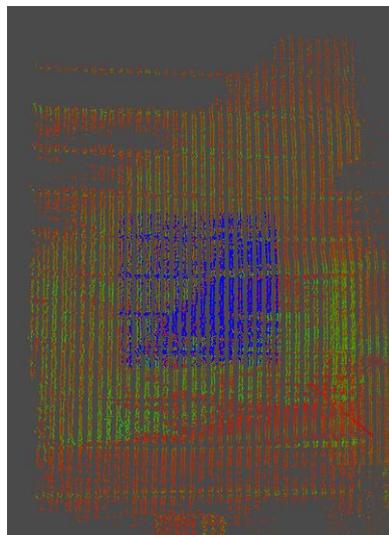


Figure 35: Reference Tile area shown in blue.

Averaging the data over 10s of sensor output resulted in the data shown above. In addition the prior knowledge about scene being aligned with the sensor’s Z axis was exploited by applying a median filter to the data to extract a surface.

From the below table the variation in the mean squared error & std. deviation of the points within the reference tile area of each sample may be seen:

	Unfiltered	Average Filtered	Average & Median Filtered
Mean Squared Error	0.000196039	2.29464E-05	1.66046E-05
Std. Deviation	0.013937871	0.00438223	0.003652621

Figure 36: Reduction of Mean squared error through filtering.

Viewing plots of this data, the damage to the tile is still undiscernible from the noise, although notably the castellations in the tile are visible as peaks running vertically through the image.

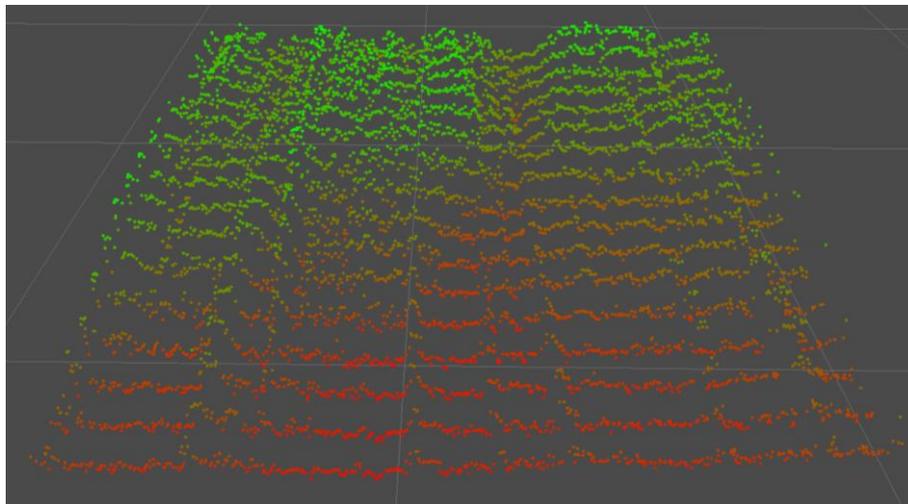


Figure 37: Rear view of centre of Tile 1, scanned with Fov3 1Hz mirror plan & coloured by depth. Due to slight gradient of tile, colour should transition linearly from Red (bottom) – Green (Top).

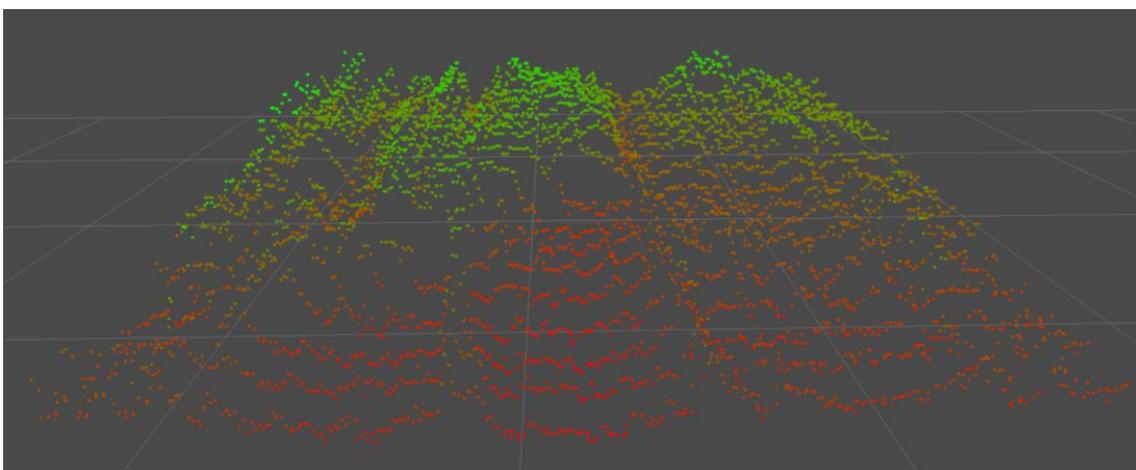


Figure 38: Rear view of centre of Tile 1, scanned with 1Hz Linear mirror plan & coloured by depth. Due to slight gradient of tile, colour should transition linearly from Red (bottom) – Green (Top).

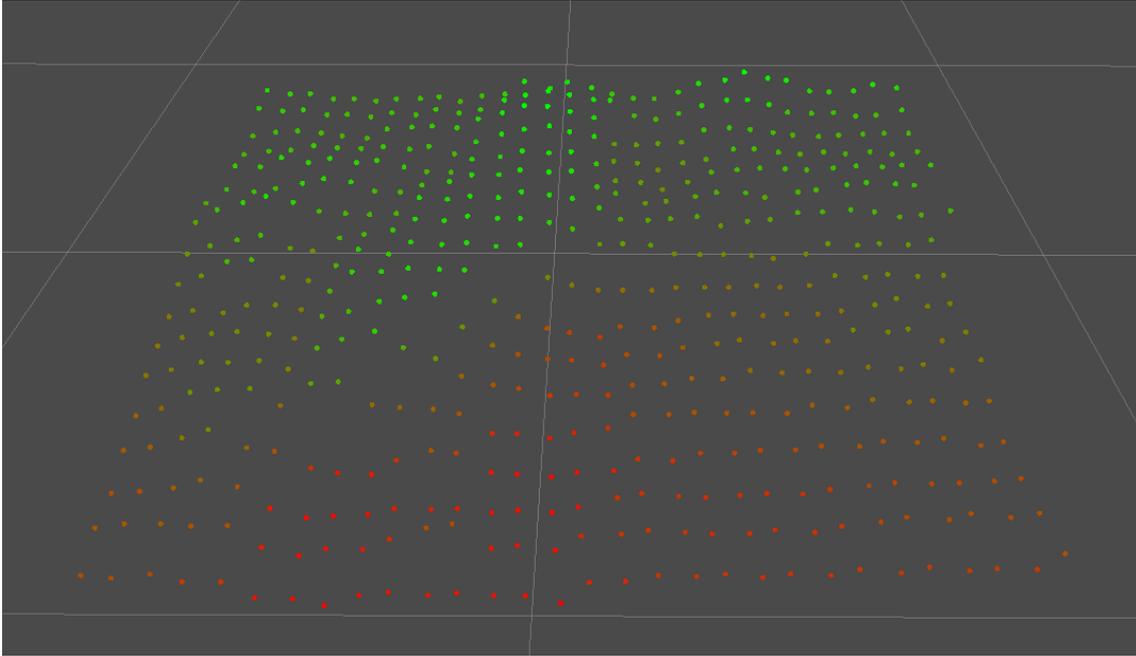
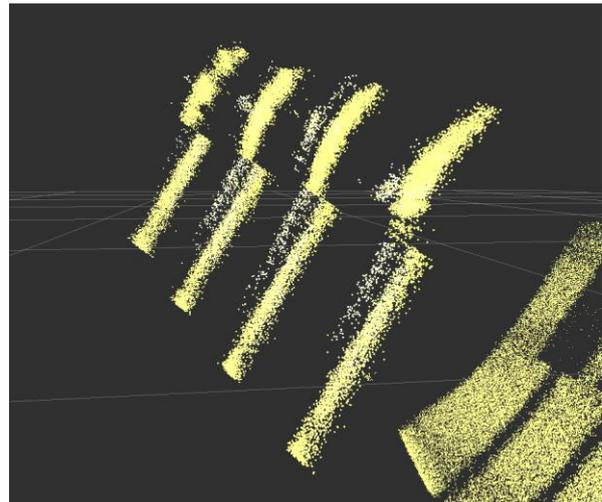
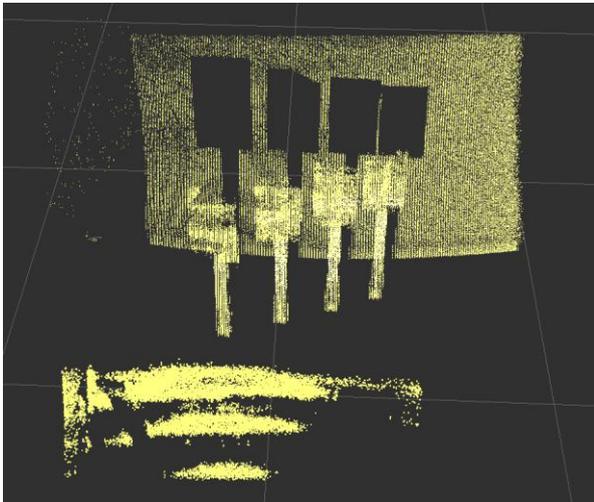


Figure 39: Rear view of centre of Tile 1, scanned with 10Hz Linear mirror plan & coloured by depth. Due to slight gradient of tile, colour should transition linearly from Red (bottom) – Green (Top).

3.1.1.5 Discussion

3.1.1.5.1 Sensor warm-up

As noted in the TACO documentation, we did observe considerable noise in the data as the sensor warmed up, however we also noted that the number of 'missing' points increased as the sensor achieved operating temperature. See comparisons below.



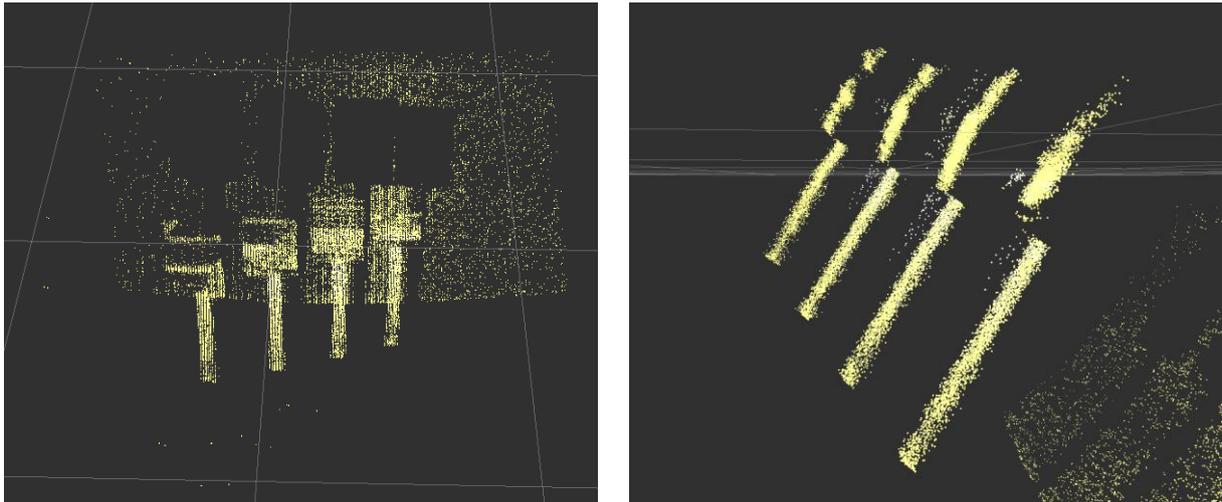


Figure 40: Sensor data during (top) and after warm-up (bottom).

3.1.1.5.2 Missing points & intensity data

It can be seen from the returned data that the tiles that are positioned further back on the test stand exhibit large areas of missing data, irrespective of the range of the test stand from the sensor. It is hypothesized that this effect is due to reflection of the laser by the metallic surfaces. See the loss of data as a tile is rotated:

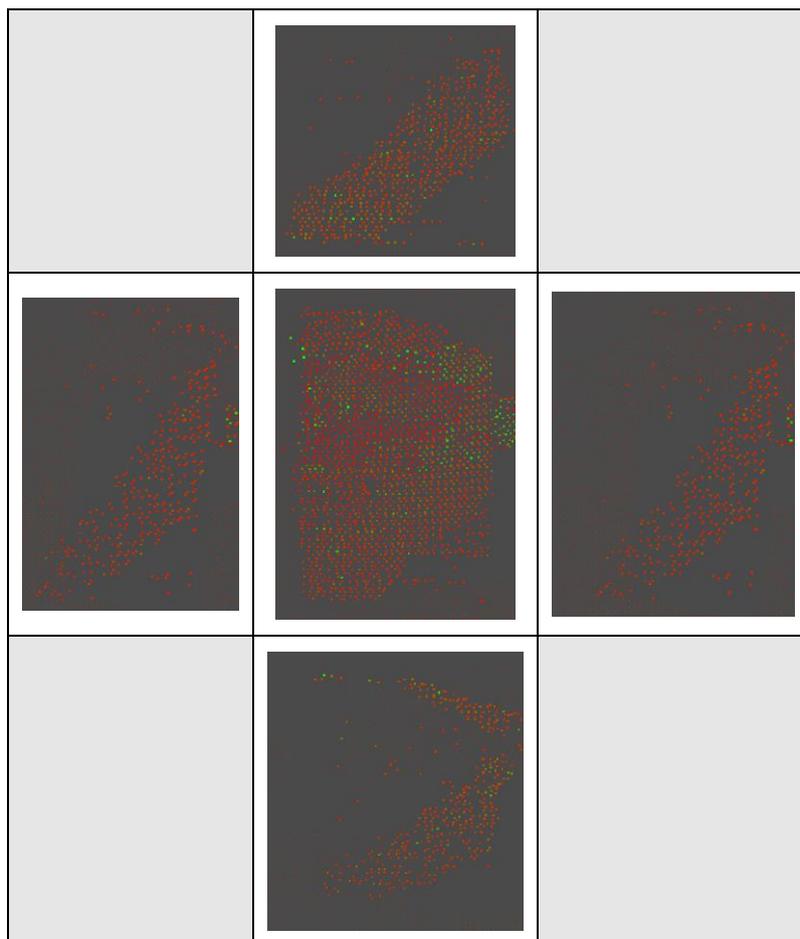


Figure 41: Tile scanned face on, rotated in TACO frame X (top & Bottom) & Y (Left & Right).

When analyzing the intensity data on the tile surfaces we noted that generally the damage to tile 1 could be observed in all positions of the test stand, however, the artifacts caused by other effects provided an equal amount of variation in the intensity data.

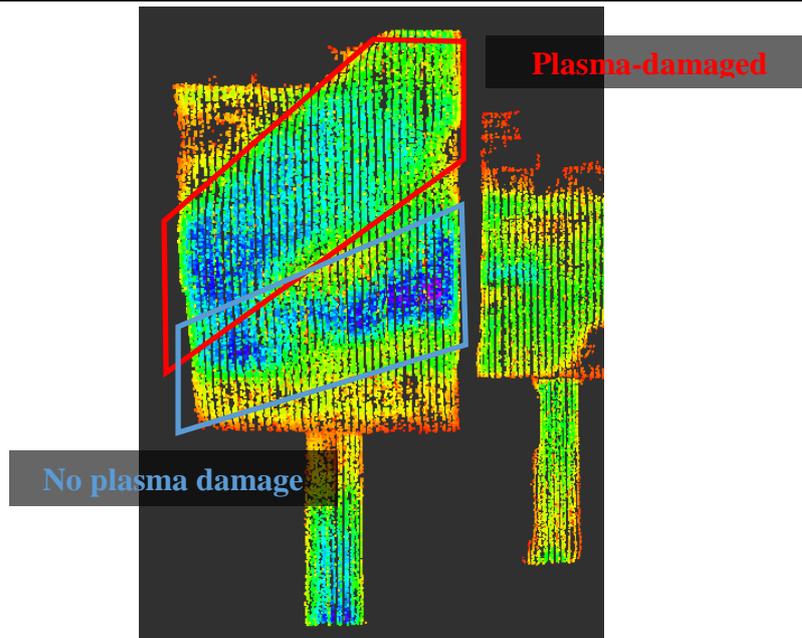


Figure 42: Damaged tile surface & artefact.

This artifact in the intensity data appeared to correspond with the area of the tile directly facing the sensor which also featured a small but visible amount of machining detail on the surface.

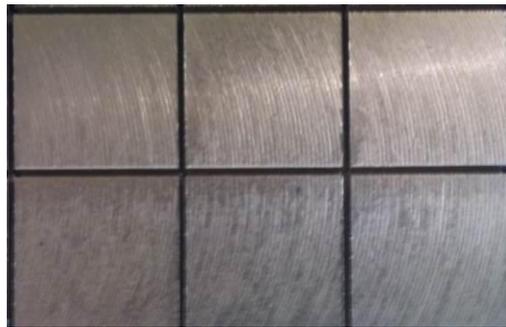


Figure 43: Tile Surface detail.

3.1.1.5.3 Non-Reflective Surfaces

To assess performance of the sensor without the issues caused by the high reflectivity of the of the surface, we scanned the tiles when covered with paper & noted a significant reduction in the surface noise, as well as the detection of the complete tile regardless of angle w.r.t. the sensor.

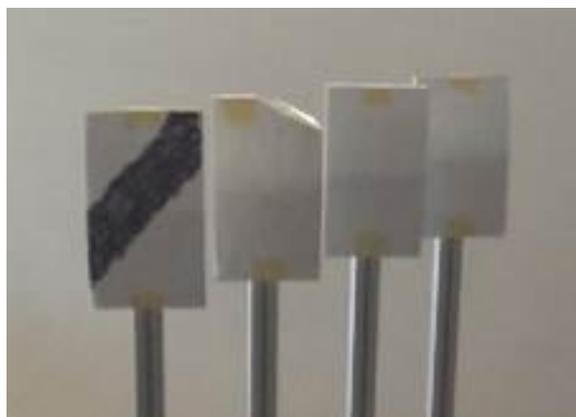


Figure 44: Covered Tiles

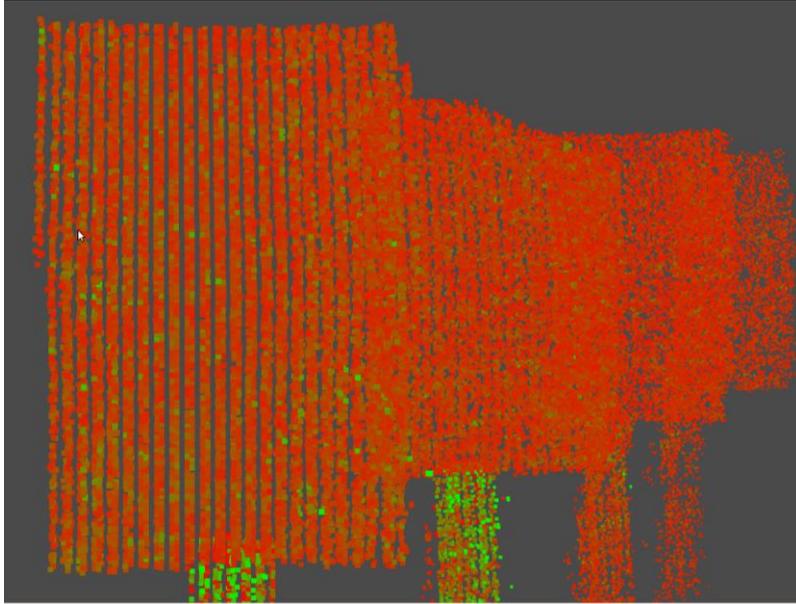


Figure 45: Covered tiles in 'Near' position scanned with 1 Hz linear mirror plan.

With the tiles covered our clustering algorithms were able to correctly identify the tiles & their centres.

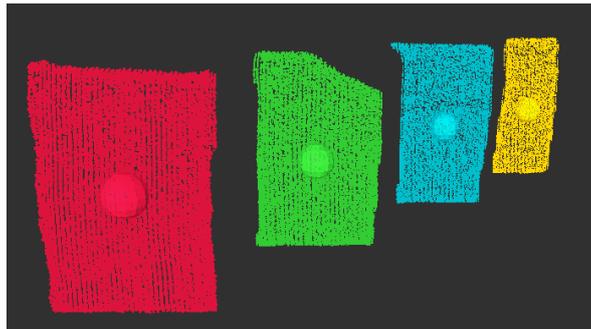


Figure 46: Clusters detected by pcl Euclidean distance grouping algorithm, with centroids.

3.1.1.5.4 X Axis oscillation

One of the most notable features that emerged from our analysis of single point noise was that there appears to be an oscillation in the X position of the data points (as shown in the plots below) that will have slightly affected the precision of our averaging filter.

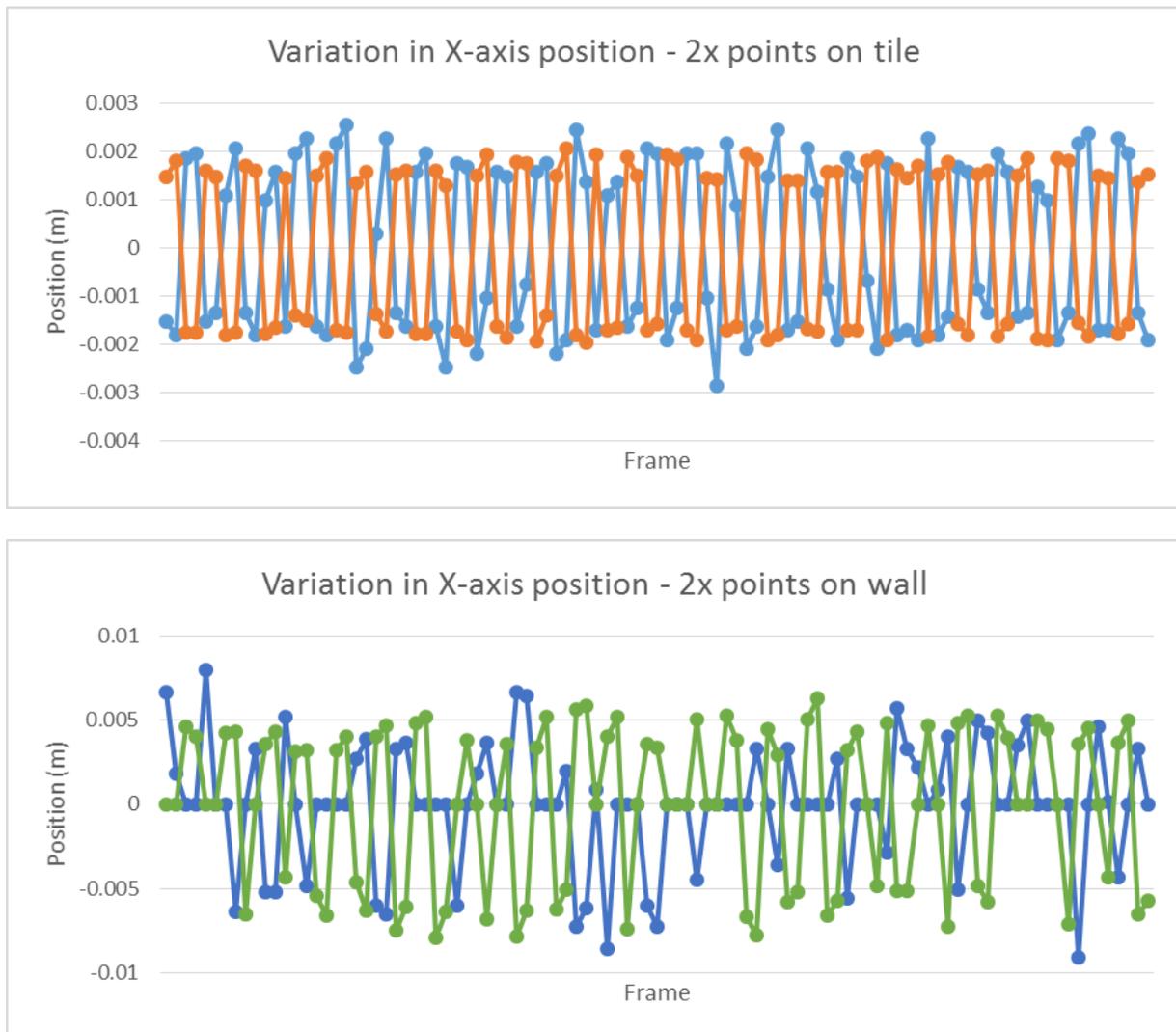


Figure 47: Oscillations along X axis of data points (0s represent no data)

3.1.1.5.5 Foveation

The foveation concept proves promising with even the limited modes made available during this trial, returning an increase in sampling density of ~50% along the Y Axis within the foveated region.

3.1.1.5.6 Conclusions

Although we have observed several issues with the Taco sensor, many of the issues are with the known issues of the high noise due to speckle & reflection/absorption of the laser. Other laser scanners have proven able to detect the intensity variations at the damaged areas of the tiles, so we are assured that this could be achieved with further work on the sensor. Whether the issues surrounding reflections off metal surfaces can be resolved is unknown to us at this time.

Should these performance issues with the laser imaging be overcome, then with the current foveation modes alone, substantial time savings could be made. To achieve a resolution of 1cm² with the 10Hz foveation mode would require that the sensor be placed approximately 2m from the scanned surface. Upon detection of damage, the ability to simply switch to the 1Hz mirror plan or a foveated view is the equivalent of moving a fixed resolution sensor head 1.8m - 1.9m (the upper limit is true if the feature can be scanned within the foveated portion of the image) towards the target. Notably, the current mirror plans would also require a rotation of the sensor by 90 Degrees.

This simpler action has two distinct advantages within a Plasma Vessel Remote handling scenario:

The time saved in approaching the feature (which would admittedly vary dependent on the configuration of the scanning robot). The robots present at JET would take at least a 2-3 minutes to perform this action & return to scanning, by contrast a rotation of the sensor could be performed in approx. 30 seconds.

The removal of the requirement to drive the remote handling systems into close proximity with the surface of the vessel (bringing the wall within 10-20cm of the sensor's optical head in the given example), would also be beneficial as it reduces risk of component damage.

With improved foveation would also come improved utility, as the sensor could be configured to either scan further from the surface or to scan at a higher density in the initial scans. It would also remove the requirement to rotate the sensor to make full effect of the increased resolution.

The current volume of the sensor is not ideal, but neither would it be problematic in current RH operating environments. The mass of the sensor however would need to be reduced in the future to allow it to be handled by the manipulators currently employed in Plasma vessel environments.

3.1.2 Augmented Reality

3.1.2.1 Set-up

The sensor was placed at a known position relative to an ABB model IRB2400-16 robot whilst the robot moved on a known trajectory emulating the placement/removal of various objects representing tiles with targeting features onto a wall representative of that found within a fusion reactor.

The Data was to be processed to estimate the position of these tiles within the scene and to generate a virtual reality display of the tile's observed pose.

3.1.2.2 Technical description of the method

The Taco Sensor was placed facing the ABB IRB2400-16 at an angle typical of either an environmentally mounted sensor, or a sensor on-board a remote platform. The ABB was placed in a scene similar to that found within a plasma vessel, with large metallic planar surfaces making up the back drop.

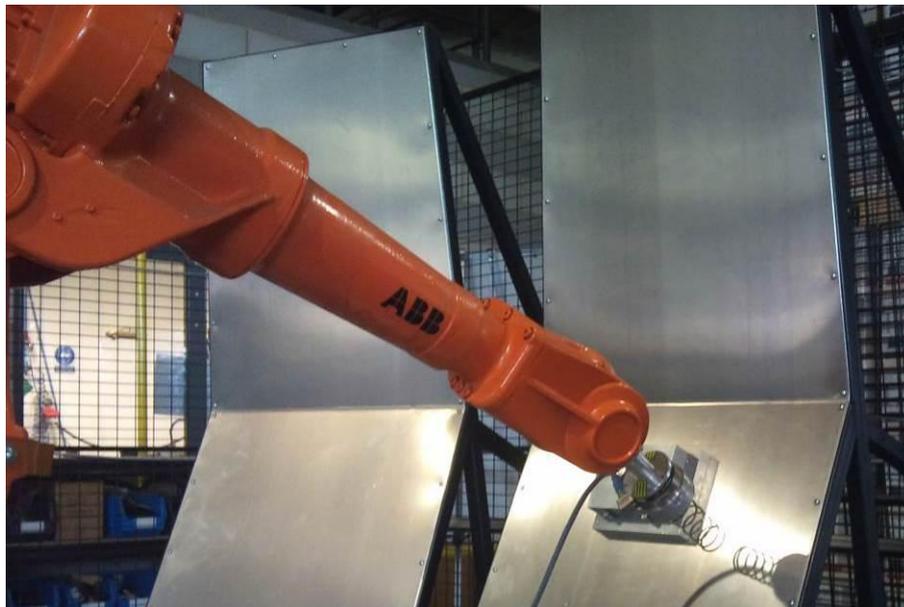


Figure 48: ABB with vessel like back-drop.

Various tiles were then placed upon the ABB tool mounting as it moved on a pre-programmed trajectory, emulating the insertion/removal of a tile at typical operations velocities.

Given the observed issues with detecting objects through intensity data (see 2.5.4) tiles had a combination of visual & physical features to supply them with a unique position & rotation within a scene. Various materials were trialed for use as markers, and the best selected for use:

- HiViz reflective strips
- White Plastic Foam Board
- Black Acrylic
- Foam Rubber
- Cotton
- White Celluloid

The TACO data was observed to detect if the markers could be identified by human operators & computational methods such as RANSAC (RANdom SAmple Consensus) and SAC-IA (SAmple Consensus - Initial Alignment) were used to locate the tile features within the scanned scene by comparison with a reference scan of the target.

3.1.2.3 Evaluation method

For all tests, visual checking of the scanned data would prove sufficient to confirm the correct operation of the sensor. For tracking the tiles, the point cloud representing the fitted target was published to a ROS topic for visual inspection & the known pose of the robot was intended for use as an absolute comparison.

3.1.2.4 Results

3.1.2.4.1 *Back drop*

As can be seen below the sensor failed to detect the metallic back drop & tile location feature. In addition it was also demonstrated that the sensor failed to detect a remote handling end effector present on the ABB at the start of the experimental set-up, nor did it detect the black paint on the ABB.

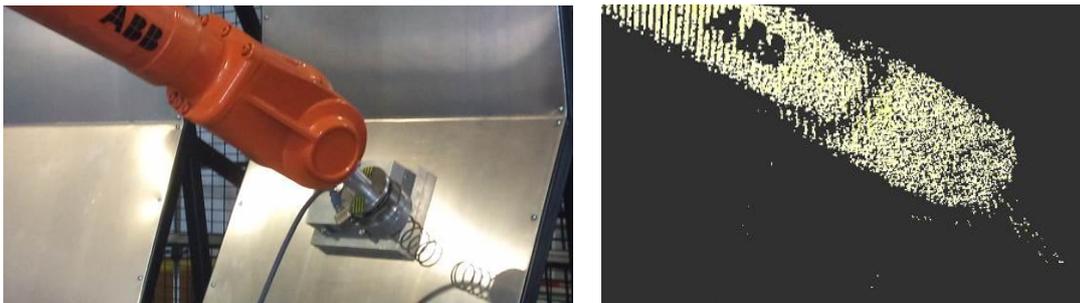


Figure 49: ABB + end effector with metallic back drop & 10Hz TACO scan of the same scene.

3.1.2.4.2 *Marker selection*

Both White plastic foam board & black acrylic were trialed as a tile material. White plastic foam board was found to be easily detectable. The black acrylic, however caused the sensor to return no data, except when aligned perpendicular to the sensor where a large amount of (noisy) data was observed. This is likely due to absorption/reflection of the laser by the acrylic.

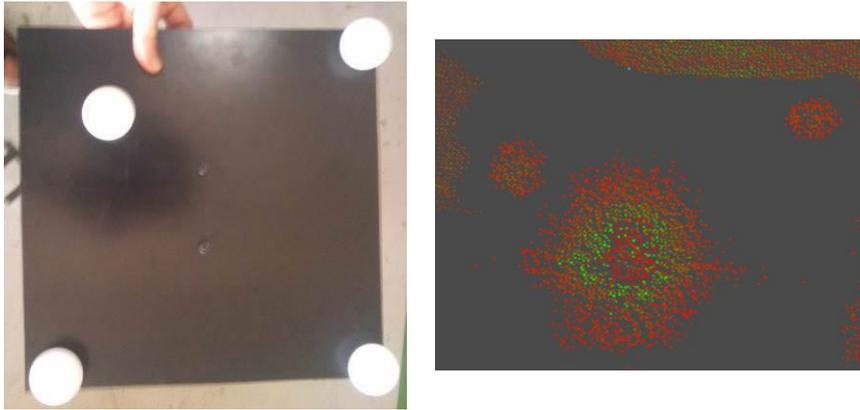


Figure 50: Black Acrylic tile & TACO image of tile with reflections.

The HiViz material was detected, although exhibited the effect of appearing deeper in the Z direction than was actually the case. It was noticeably brighter in the intensity data.

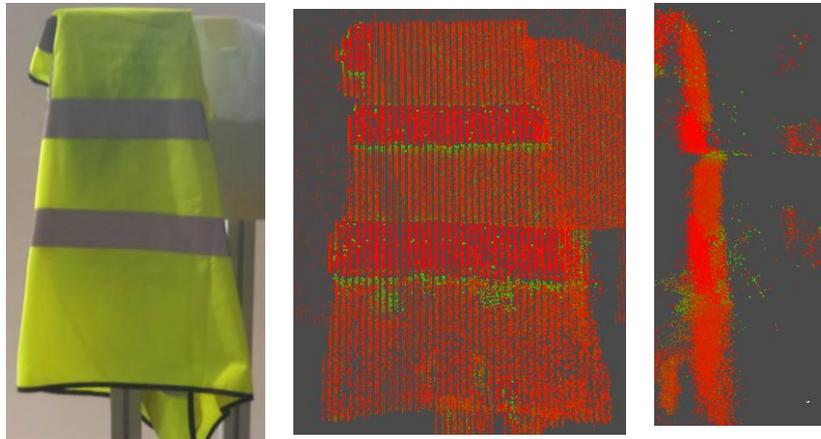


Figure 51: HiViz material on jacket, TACO scan of jacket face & side on.

3 spheres (celluloid, cotton & foam rubber) were scanned to assess their utility as markers. Only the celluloid sphere did not appear with a noticeable cloud of noise located in front of it.

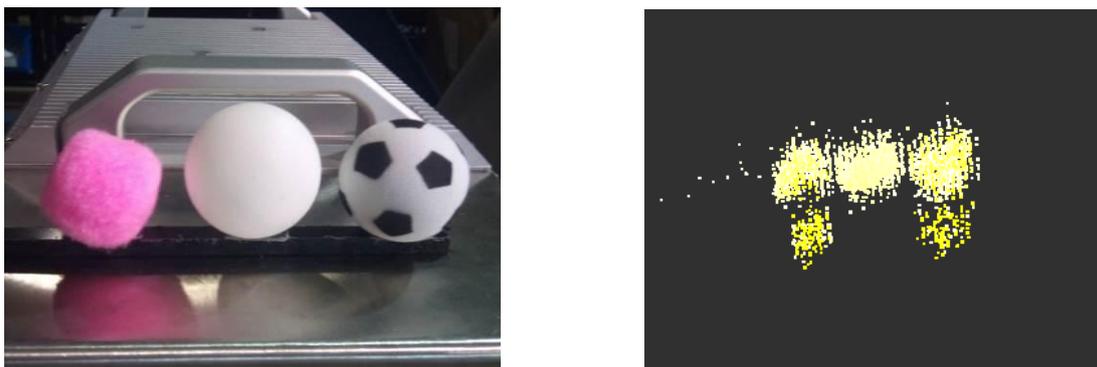


Figure 52: Cotton, Celluloid & Foam rubber spheres & top down TACO scan.

A feature constructed of the White plastic foam board was manufactured representing the typical size of a feature that could be machined into a Tile carrier typically used at JET, however when scanned was almost undetectable from the surface noise.

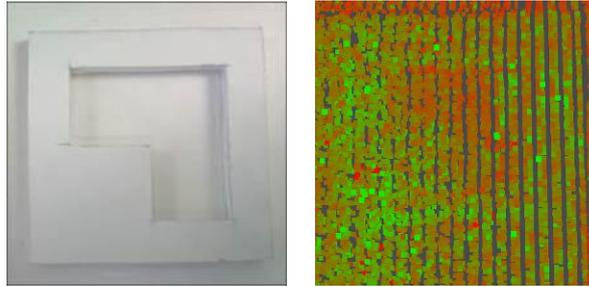


Figure 53: White plastic foam board target & Taco scan.

Therefore a larger feature such as could be deployed on the larger components at ITER was used.



Figure 54: Revised White plastic foam board target.

3.1.2.4.3 Marker detection

Two tiles with markers were constructed. The first from White plastic foam board with the L shaped marker constructed from the same material. The second consisted of a black acrylic tile with celluloid spheres on the surface (Figure 50). The use of the black acrylic was to highlight the sphere within the scene by providing no background.

Sphere tracking was attempted via RANSAC, however the small size of the spheres combined with the amount of noise on the sphere's surface meant that tracking was not possible (both sphere & circle fit methods were attempted with no success).

SAC-IA tracking was attempted on both tiles. During acquisition of reference tile data, an odd warping effect was noted, where the L shaped reference tile was 'folded over'.

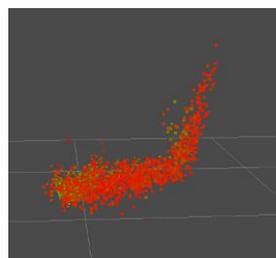


Figure 55: Warping of target when viewed side-on.

Both tiles were tracked against the reference data, however the amount of noise in the image caused the feature detection algorithms to struggle.

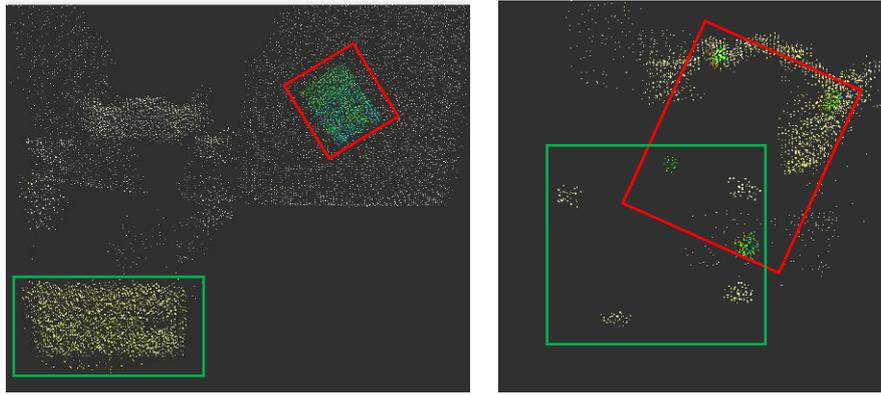


Figure 56: Tiles (highlighted in green) & computationally identified positions (highlighted in red).

Another notable effect that was observed when tracking the markers in 1Hz mode was the warping of the data due to the motion of the robot. This caused the straight surfaces of the tiles to appear warped, due to the robot assuming a new pose between the start & finish of the scan.

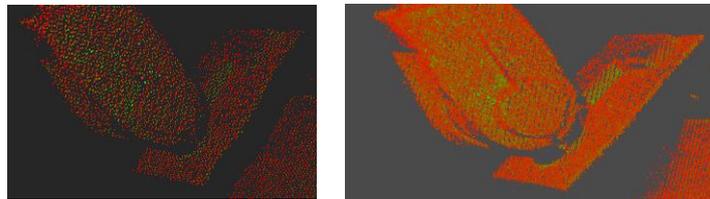


Figure 57: ABB scanned with 10Hz & 1Hz mirror plans, note the increased angle in 1Hz scan due to robot motion.

3.1.2.5 Discussion

Ultimately the current feature set of the TACO sensor does not suit itself to this type of activity. Some of the tile features are clearly visible in the returned data, but the noise in the image prevented the trialed fitting methods from working correctly & would cause inaccuracy even in a functioning method. However as discussed in 3.1.1.5 many of these problems are with the noise in the data and the reflectance of the laser by metallic surfaces, which might be overcome in the future.

Another issue with the sensor as it stands, is the slow frame rate that must be used to allow foveation, which produces warping in the image when moving objects are observed. The ability to foveate at high speed would rectify this.

A desirable feature would be for the TACO sensor to be able to auto correct the point cloud data based on velocity data provided from an external source, thus allowing the sensor to be operated from a moving base.

3.2 Public Safety (SHADOW)

The goal of this use case is to evaluate the 3D data provided by the TACO sensor for the task of object recognition and tracking for visual servoing with a performance comparison against the Kinect sensor. This use case highlights the use of foveation capabilities of the sensor allowing for detection of smaller objects to track compared to commercial sensors like the Kinect. Additionally, we showcase the robustness of the TACO sensor to different lighting conditions by tracking objects in outdoor like environments under direct sunlight and in no light.

3.2.1 Visual serving

3.2.1.1 Set-up

Workspace prepared with a central table, where the suitcase would be. At one end of the table is the arm and hand, ready to servo and accept the swab. The swab itself starts in the middle of the table. On a separate bench next to this table the TACO sensor and Kinect are setup with a full view of the scene. Also here are a set of powerful, sunlight like producing lights. Mounted on the lights is a basic web camera used to record a video stream for each run.

The entire setup is surrounded by a black, light absorbing curtain for safety and to control the lighting environment.

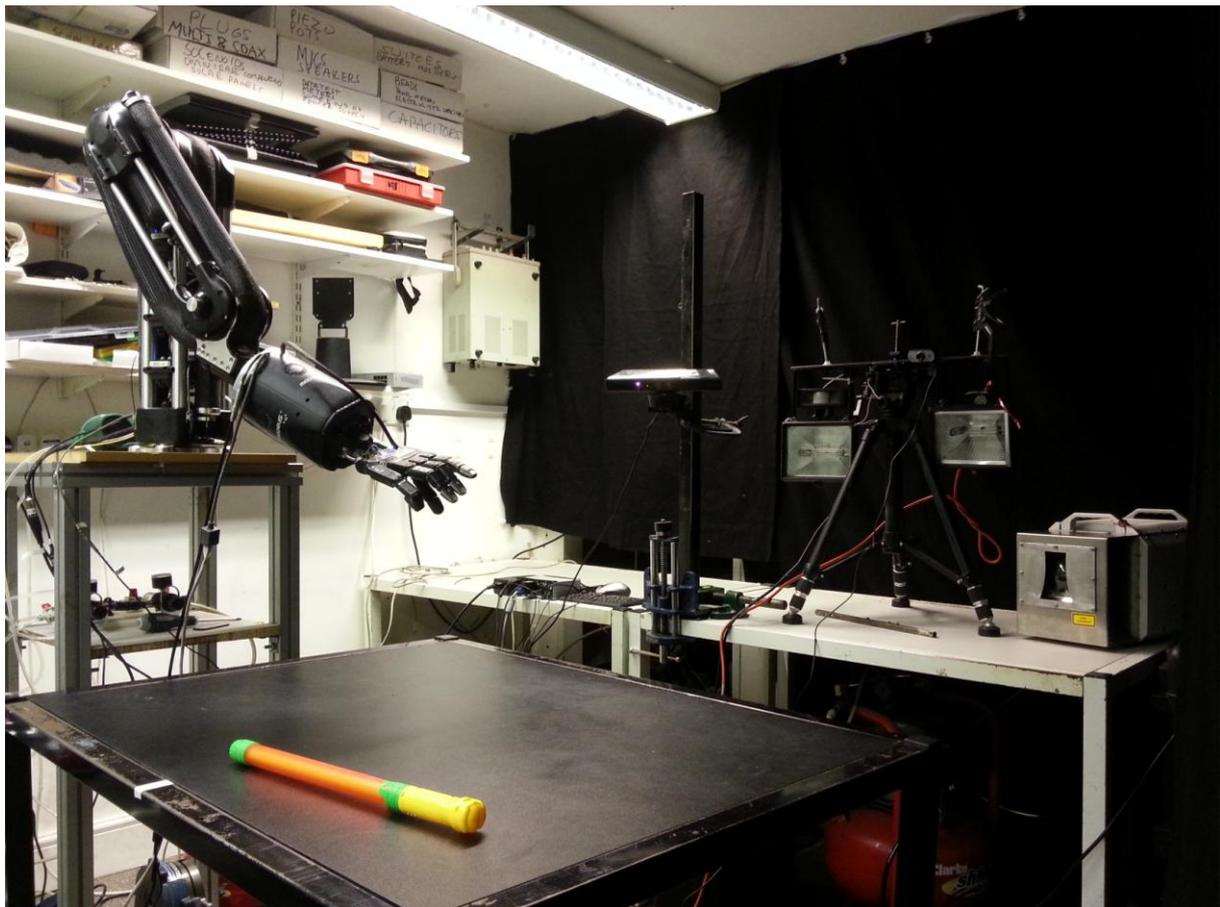


Figure 58: Public safety test case setup

3.2.1.2 Technical description of the method

The same basic use case is run through under a number of lightning conditions; normal indoor lighting, darkness (no lights) and strong outdoor lighting (simulated using bulbs) for both the TACO sensor and Kinect sensor.

1. Swab starts at rest in middle of the table.
2. TACO sensor started in 1Hz mode, high density.
3. Initial object to track is segmented from the center of the scene (the swab). This cloud is used to seed and start the tracker,
4. TACO sensor switched to 10Hz mode (lower density) for tracking.

5. Short period with no motion.
6. Human operator grabs the swab and moves it through a series of motions.
7. Swab return to rest on the table.

Steps 2 and 3 are skipped for Kinect runs as it has no such feature.

This whole case is recorded to a ROS bag file containing the following topic data:

- **/gscam/image_raw** (sensor_msgs/Image) – RGB web cam recording video of the use case run. Used for reference, not analysis.
 - **/sr_pcl_tracker/cloud_downsampled/points** (sensor_msgs/PointCloud2) – The target cloud. Pointcloud used by the tracker for finding the object. This is a processed version of the TACO sensor cloud.
 - **/sr_pcl_tracker/result/points** (sensor_msgs/PointCloud2) – The cloud for the object returned by the tracker aligned to the current target cloud. Where the tracker thinks the object is.
 - **/sr_pcl_tracker/result/pose** (geometry_msgs/PoseStamped) – 6DOF pose calculated by the tracker for the result cloud. Used to drive the servoing.
1. **/taco_sensor/combined** (sensor_msgs/PointCloud2) – Combined cloud with both foveated and unfoveated frames.
 2. **/taco_sensor/foveated** (sensor_msgs/PointCloud2) – Unfoveated raw cloud from the TACO sensor.
 3. **/taco_sensor/unfoveated** (sensor_msgs/PointCloud2) – Foveated raw cloud from the TACO sensor.

3.2.1.3 Evaluation method

Analysis was performed on the bag files recorded as above.

We consider the cloud processed by the tracker (**/sr_pcl_tracker/cloud_downsampled/points**) as the dataset for each frame, this is the cloud received from the sensor after some pre-processing has been applied - filter on the z axis and down sampling. This pre-processing is the same for both sensors. Note that this results in a different size cloud for different frames.

For each point we can assign a True value for, is part of the object, and a False value for, is not part of the object. Therefore the null hypothesis is: point is not part of the object.

We can then apply this contingency table to each frame, counting the number of true and false points:

		Tracker Result	
		True	False
Expert Result (Human annotated)	True	TP – True Positive	FN – False Negative
	False	FP – False Positive	TN – True Negative

Initial processing via python script to create CSV file of bag time, total_points (the target cloud). Bags are recorded with high res (nanosecond) time, with interleaved messages for the different message types as they are published. The pre-processing script therefore fills in missing values for the total using the last value seen, I.E. the last value published into the ROS system at that time. We also use this approach to fill in those values for annotated rows.

This CSV was then loaded into a spreadsheet.

Each bag file was played back and examined using Rviz, stopping every couple of seconds through the scene to annotate a frame. Recording this data for each annotation:

bag_time: Time we stopped at in the bag.

notes: Text notes on any interesting features of that frame.

true_points: Human expert counted number of points in the object.

tp_points: True positive points, number of points in the object the tracker correctly identified as part of the object.

fp_points: False positive points. Points identified by the tracker as in the object that actually lie outside the object.

The counting was done using the rviz selection tools on the cloud.

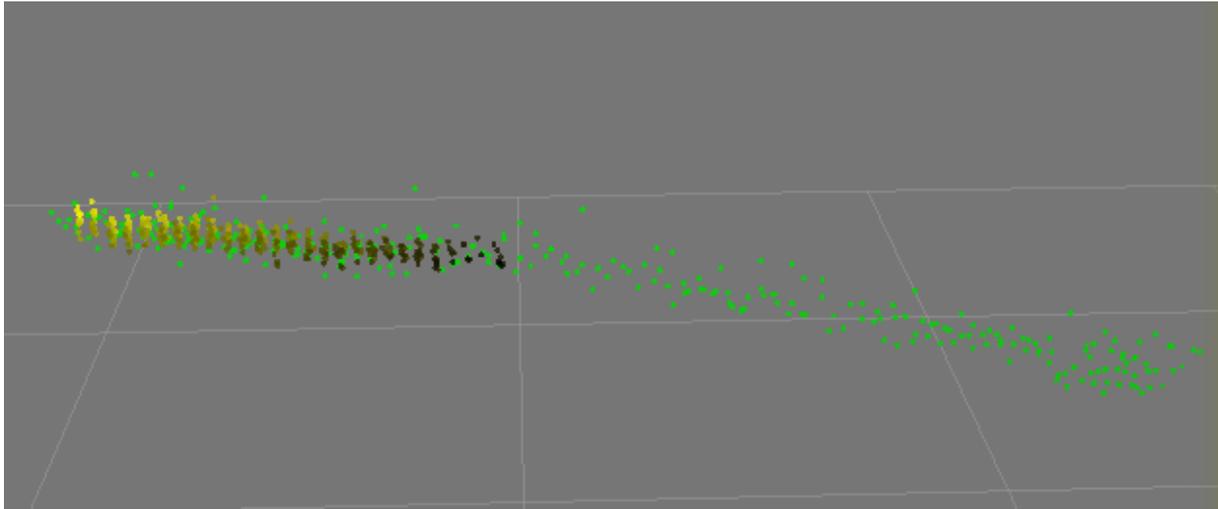


Figure 59: Example object cloud

The green cloud is the target cloud for the object its self, we count this to get *true_points*.

The yellow cloud is the tracked object returned by the tracker. The tracker works by being seeded with an initial cloud to track and then once running returns that original cloud transformed to best fit the target cloud. It does not however actually give us the points in the target cloud. Additionally for the TACO sensor the density of these two clouds are different because of the change from 1Hz to 10Hz scanning (described below). For these two reasons to get the value for *tp_points* we count the number green points inside the area given by the yellow cloud. Likewise we count any green points not part of the object but inside the yellow points as *fp_points*. This gives us comparable values against the same cloud.

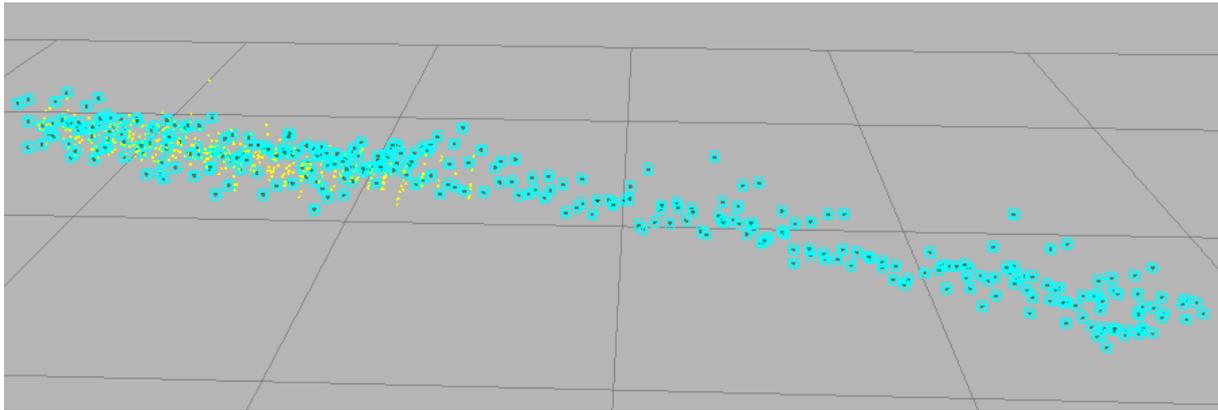


Figure 61: Counting true points

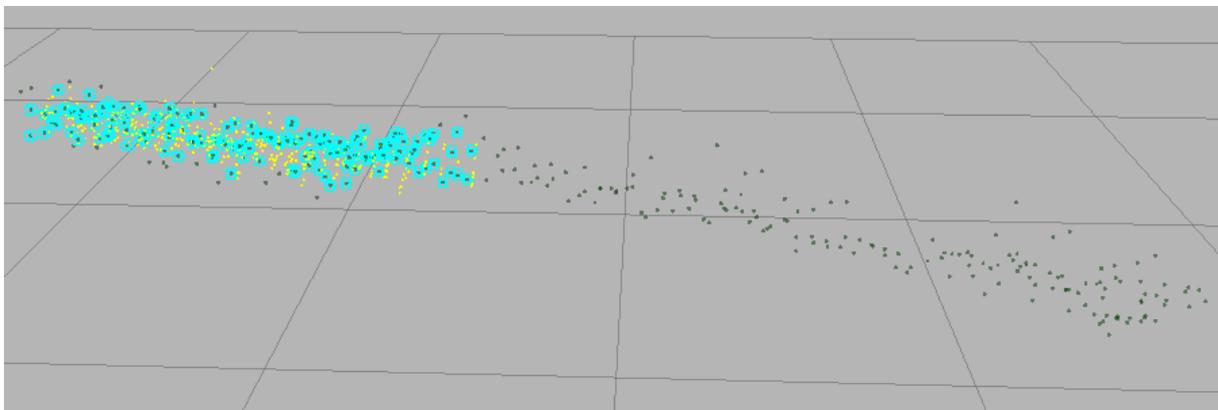


Figure 60: Counting true positive points

The annotated data was then processed in a spreadsheet to produce the following extra fields.

fn_points: False negative points, points the tracker labeled as not in the object that actually are. Calculated as $true_points - tp_points$

target_points: Number of points in the target processing cloud, looked up from the processed bag file.

false_points: The actual number of points not in the object, expert counted. Calculated: $target_points - true_points$

tn_points: True negative points. Points the tracker and the expert agree are not in the object. Calculated: $false_points - fp_points$

accuracy: A simplistic measure of the accuracy of the tracker. What percentage of the true points did the tracker identify as part of the object correctly. $tp_points / true_points$

chi_squared: The Pearson's Chi Squared test value for the contingency table for this frame. Note that this value is specific to the row and its sampled size. Higher values indicate higher correlation.

p_value: Probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Calculated from the chi-square value by comparison to the normal distribution. This gives us a measure of the data in this frame is significant using the standard threshold of 0.05.

phi: Phi coefficient (mean square contingency coefficient) to give us another measure of the accuracy of the sensor, the correlation of the sensors result to the experts result. Calculated from the chi-square value, normalised by sample size so is comparable across frames and sensors.

3.2.1.3.1 TACO Sensor Density change

The TACO sensor was run at 1Hz, high density but too low frame rate for tracking, for the initial tracker seeding, which is finding a target cloud to track and giving that cloud to the tracker. Then for the actual tracking the rate is switched to 10Hz, lower density but fast enough for tracking.

Unfortunately the tracker returns this original seed cloud, transformed to its new position in the frame rather than the actual points matched. This means it maintains its original high density. This is clear in the following images, where the blue cloud is the sensor cloud and the yellow the tracker result cloud.

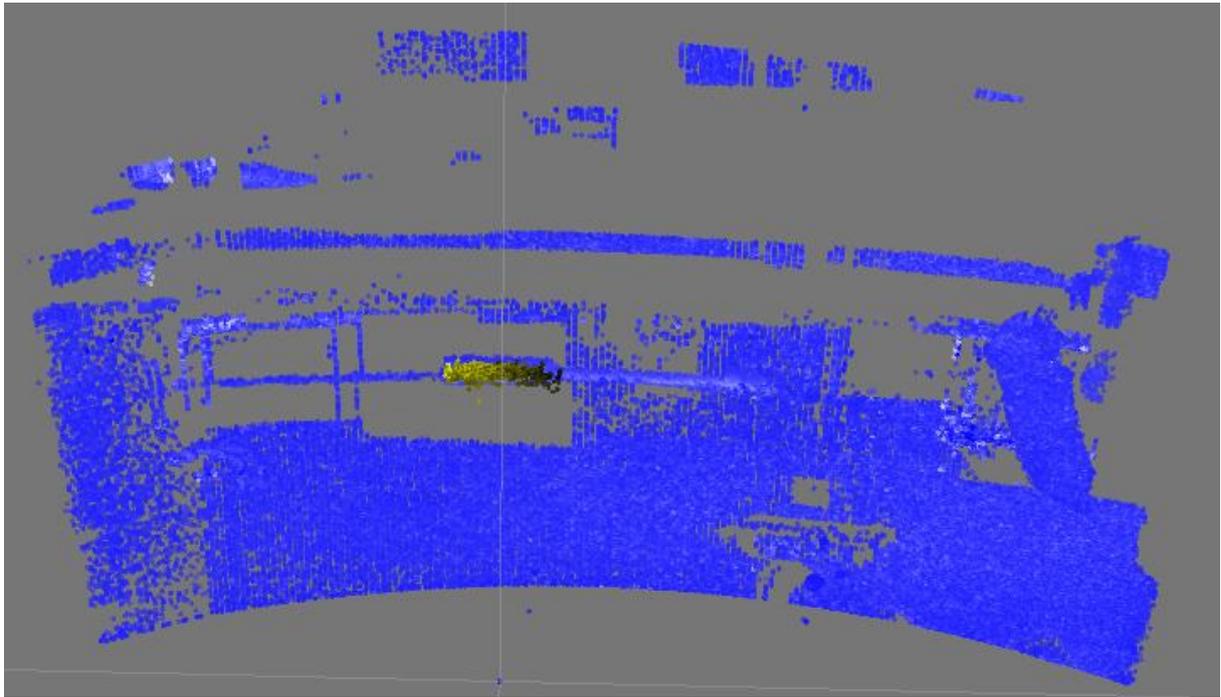


Figure 62: 1Hz high density TACO cloud and high density tracked object

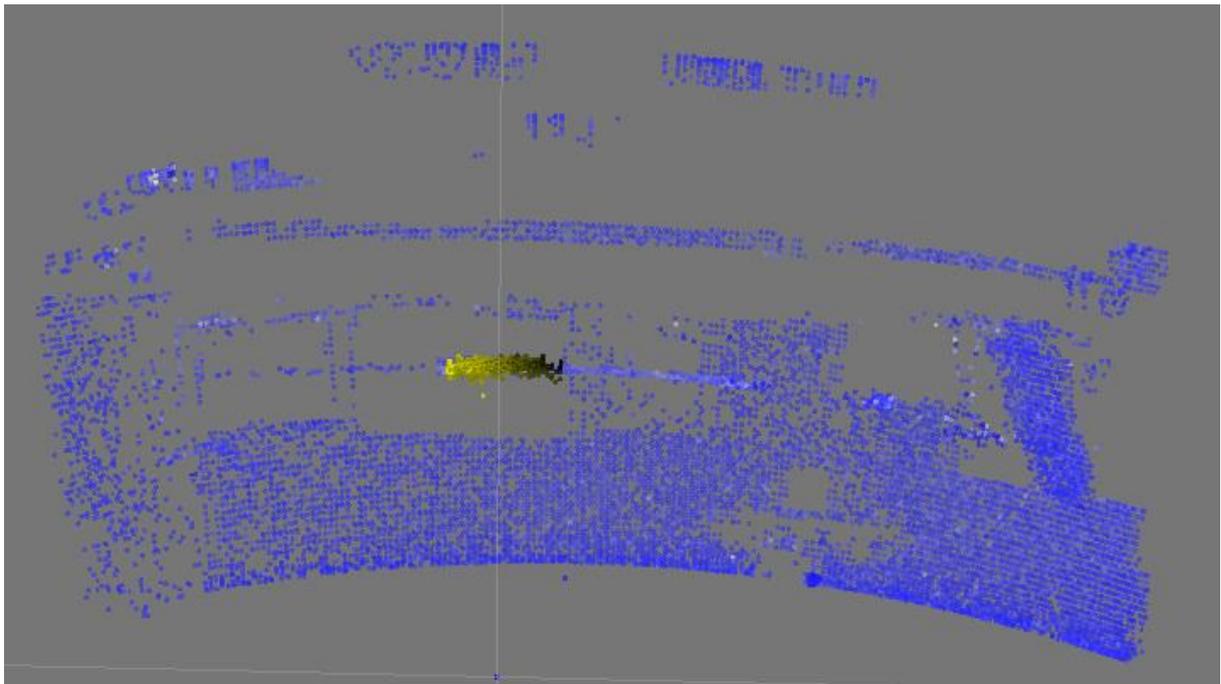


Figure 63: 10Hz low density TACO cloud and high density tracked object

3.2.1.4 Results

All annotated frames had a p value of 0.0 to 10 decimal places so we consider them all to be statistically significant with respect to the null hypothesis.

	Average accuracy in lighting conditions		
Sensor	Normal	Strong	Darkness
TACO	48.41%	52.99%	56.21%
Kinect	77.13%	0.00%	80.89%

	Average phi in lighting conditions		
Sensor	Normal	Strong	Darkness
TACO	14.87	19.56	26.28
Kinect	129.05	0.00	136.80

3.2.1.5 Discussion

Under normal and dark lightning conditions the TACO sensor performed worse than the kinect sensor. Foveation appears to give no advantage in this case.

In strong outdoor light the TACO sensor outperformed the kinect which failed to produce a usable cloud. The TACO technology is therefore a good choice in outdoor work.

While foveation was expected to increase the accuracy of the tracking this was not seen in practice due to other problems with the sensor. The known inaccuracies of distance measuring, noise in the cloud, lower resolution and curvature of the cloud would appear to all be more significant in reducing accuracy than any increase from foveation.

The true negative count is always very high as the object is a small number of points compared to the target cloud.

We don't see false negative points due the tracker using a target cloud slightly smaller than the tracked object and the object cloud tending to be out in space so there are no non object points near enough.

3.3 Home Grasping (TUW)

3.3.1 Self-localisation

This use case was discarded due to the little time available and in order to spend more time with the object detection use case that is more interesting in order to showcase the foveation capabilities of the sensor.

3.3.2 Object detection

The goal of this use case is to evaluate the 3D data provided by the TACO sensor for the task of object recognition as well as a performance comparison with data provided by the Kinect sensor. This use case is of special importance to highlight the foveation capabilities of the sensor allowing for instance a mobile platform to detect objects at larger distances compared to commercial sensors like the Kinect. Additionally, we showcase the robustness of the TACO sensor to different lighting conditions by detecting objects under direct sunlight.

3.3.2.1 Set-up

The use case is designed as follows: Different objects (one at a time) are placed in front of the sensors at different distances (from 1 to 2.5 meters with approximately half meter interval). For the TACO sensor, one unfoveated (1Hz linear trajectory) and one foveated frame are recorded. A 3D model of the objects to be detected is available and used to train the object detectors (see Figure 64).

Please note, that the object detectors are trained only once and the same detector is used for both Kinect and TACO data. There are 4 objects to be recognized: a cylinder, a mug, a spray bottle and a tetra pack. The 3D point clouds were generated by registering and fusing several Kinect views around the object.

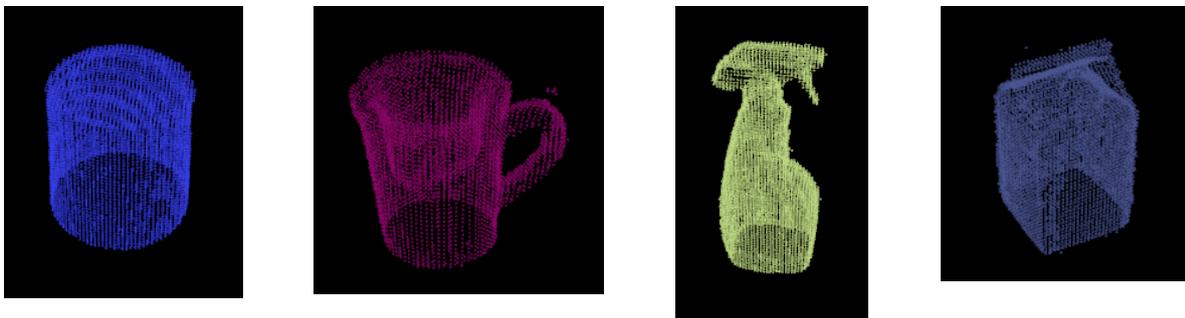


Figure 64: 3D models used to trained the object detectors

The TACO sensor is mounted on a mobile platform providing a tilting mechanism on which the TACO sensor is located, allowing it to look directly at the ground where the objects are placed. Depending on the distance, the tilting angle is adapted using a linear actuator so that the objects are always visible. The Kinect sensor is mounted on top of the TACO sensor thus providing similar viewpoints. Figure 65 depicts the setup.

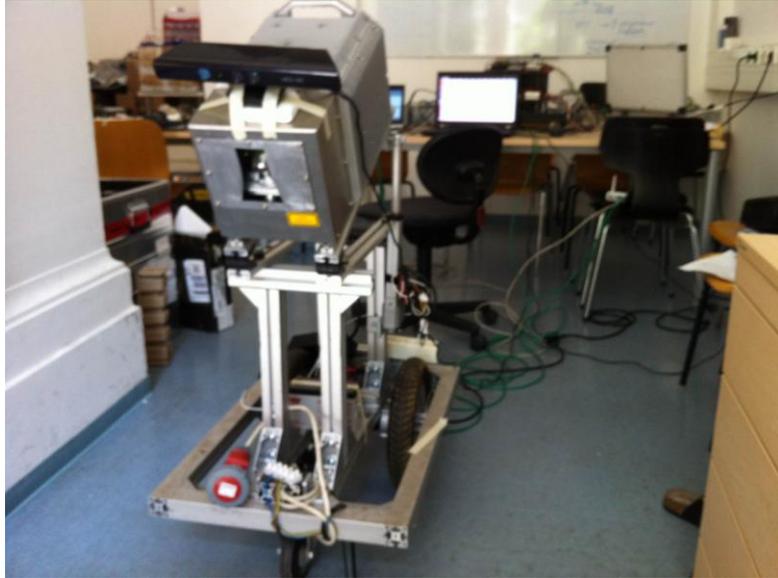


Figure 65: TUV setup: Kinect mounted on top of the TACO sensor which is mounted on a mobile platform with a tilting unit. The mobile platform allows us to capture point clouds of the objects at different distances.

3.3.2.2 Artifacts with the TACO sensor

We observed several artifacts that have complicated the deployment of this use case. One of the most severe artifacts that we observed was caused by large flat surfaces being reported as wavy surfaces in the point cloud of the TACO sensor (see Figure 66). Such an artifact prevented us from using 10Hz data in our experiments as the recognition method assumes a planar surface in order to segment the objects of interest. The artifact is less visible in 1Hz linear and foveated trajectories thus allowing us to use this data for the evaluation.

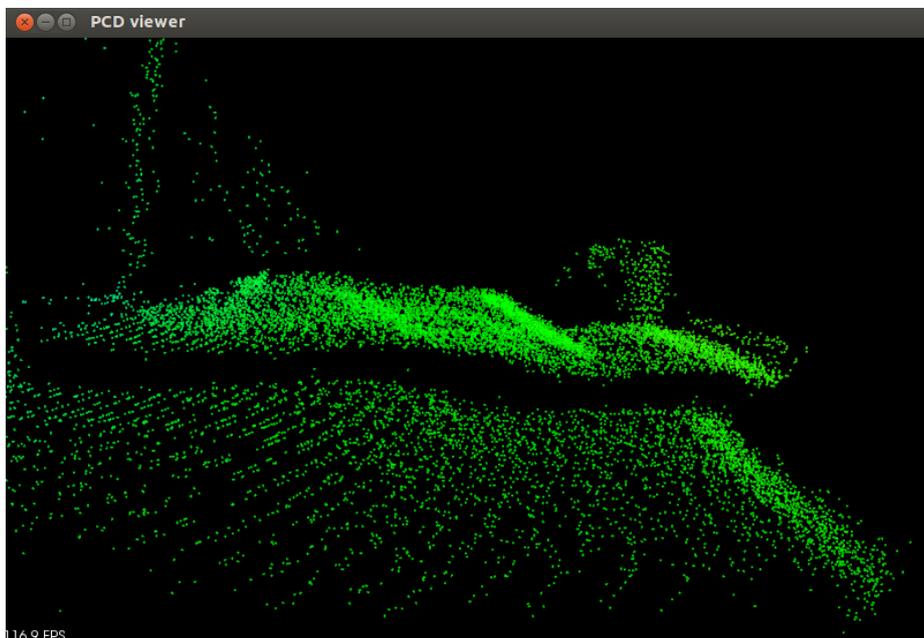


Figure 66: Flat surfaces are reported as wavy surfaces

A second artifact was noted due to the miss-calibration between the different channels in the TACO sensor. This was causing "shadows" in front of the actual surfaces. We observed that this effect is severe when the laser power is set to a high value (> 150). Reducing the power of the laser caused the artifact to disappear as all measurements are registered to the most sensitive channel. A proper calibration between different channels would be desired and would improve the data quality.

Unfortunately, the calibration process is a time consuming trial and error process that we were not able to perform due to the limited amount of time available for the tests. Figure 67 shows this artifact in front of the mug where the planar surface is reported twice at different locations. The power of the laser is also related to the artifacts previously reported at 0.7m. For instance, with current > 150, the artifacts are visible. Reducing the power of the laser causes far away surfaces to be undetected (especially for surfaces with a small angle of incidence --- almost parallel to the laser ray).

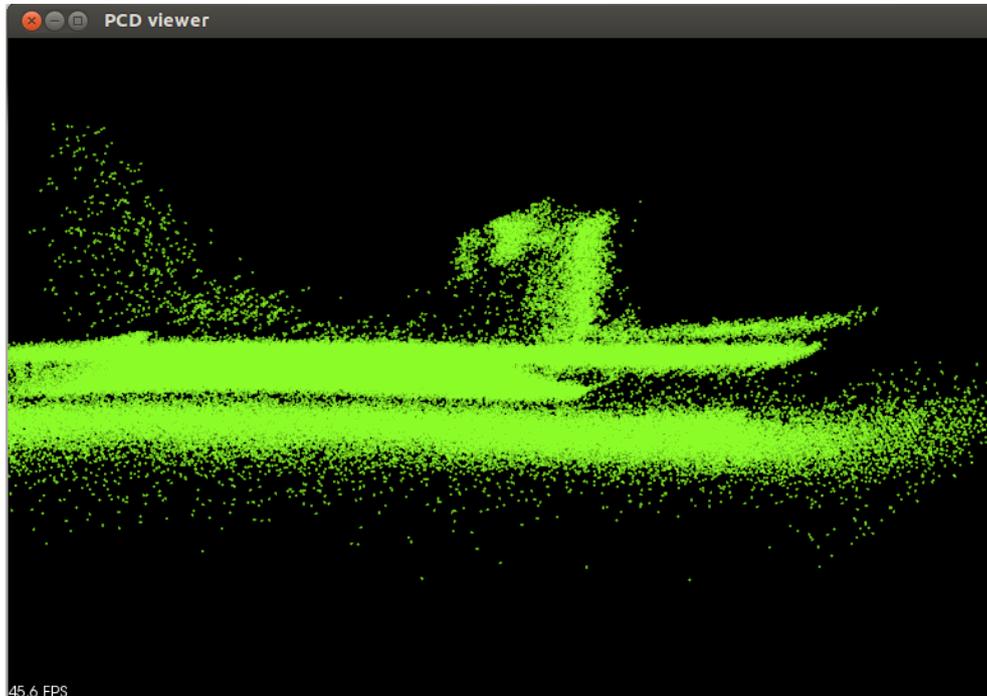


Figure 67: 1Hz trajectories do not present wavy planar surfaces but with high laser power, “shadow” effects can be observed (see text for a more detailed explanation).

3.3.2.3 Technical description of the method

After a first visual analysis of the TACO data and taking into account the amount of observed noise in clouds, we decide to perform object recognition applying global methods instead of local methods. Global methods characterize themselves by requiring segmentation prior to recognition; once the object of interest has been segmented out from the background, a global descriptor (a histogram encoding the global geometrical features of the point cloud) is computed and matched against the training dataset. Because, global geometric features are included, these methods are usually much more efficient under the presence of noise. For classification, we use a 1-NN classifier and therefore assign to each segmented object the identifier of the closest object in the training set. Concretely, we use the following global descriptors:

- Viewpoint Feature Histogram (VFH): computes angle distributions between the surface’s normals. See [1] for more details.
- Ensemble of Shape Functions (ESF): computes an ensemble of shape distributions (i.e., distance between pairs of points). Does not require the computation of surface’s normals. See [2] for more details.

During training, the 3D models representing the objects of interest are rendered from different viewpoints in order to simulate the 2.5D properties of a scan obtained from a single viewpoint. In other words, self-occluded points are computed and discarded before computing the descriptor. During these experiments, we rendered 36 different views of the objects. Figure 68 shows how these generated partial views look like for one of the objects used in the experiments. Using a simple 1-NN classifier allows us to easily assess the quality of the data provided by both the Kinect and the TACO

sensor compared to the training data (which is almost noiseless and free of artifacts because of the virtual rendering).

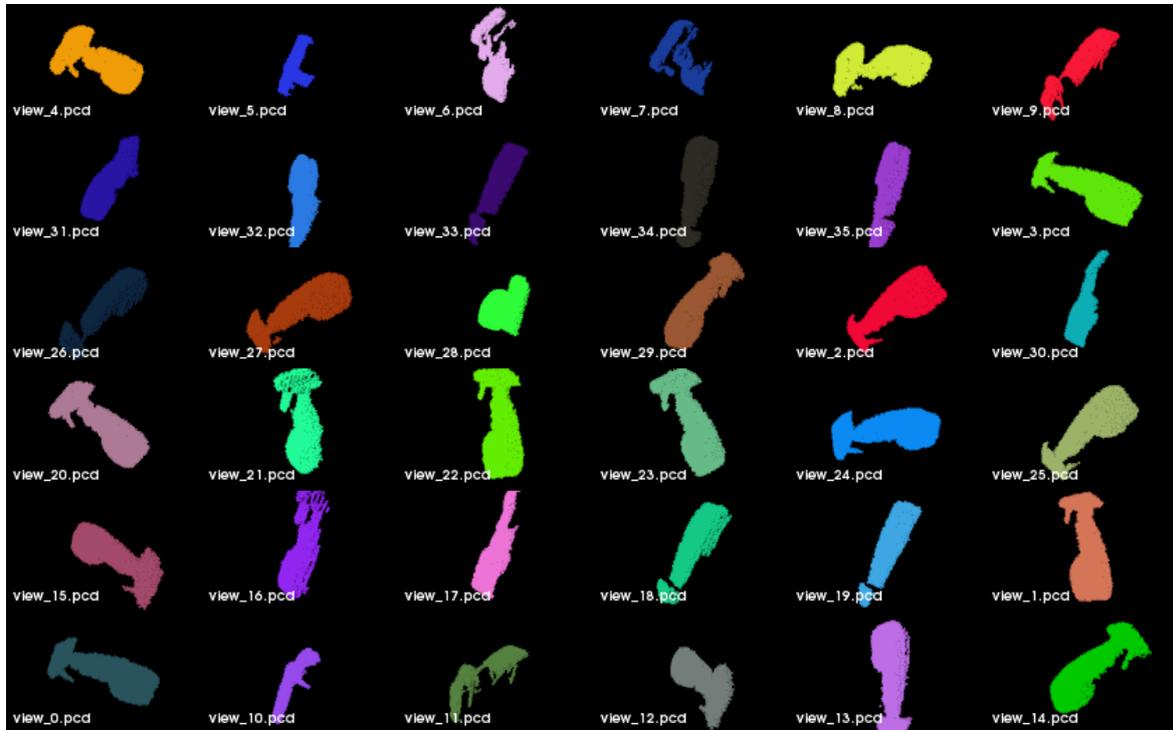


Figure 68: Partial views used to train the object detectors. The views are obtained by rendering the 3D model representing the object of interest (i.e. the spray bottle)

3.3.2.4 Results

The evaluation of the results is summarized in the following tables. In Table 4 whenever the object is correctly recognized, we report 1, 0 otherwise. Results are reported for the 4 objects to be recognized (for Kinect and TACO data; 1Hz unfoveated and 1Hz foveated for the latter) and the different global descriptors. Distance to the sensor increases from 1m (Scene 1) to 2.5m (Scene 4) with a half meter interval. Table 2 summarizes the recognition results grouped by distance for the different types of data and descriptors. Similarly, Table 3 reports results grouped by object type.

	Descriptor	1m	1.5m	2m	2.5m	Average
TACO (1Hz unfoveated)	<i>ESF</i>	75%	50%	50%	50%	56.25%
	<i>VFH</i>	75%	50%	50%	50%	56.25%
TACO (1Hz foveated)	<i>ESF</i>	50%	25%	75%	25%	43.75
	<i>VFH</i>	75%	50%	25%	50%	50%
Kinect	<i>ESF</i>	100%	75%	50%	50%	68.75%
	<i>VFH</i>	100%	75%	100%	75%	87.5%

Table 2: Recognition results grouped by distance for Kinect and TACO (1Hz foveated and unfoveated) data

	Descriptor	mug	Spray bottle	Tetra pack	Cylinder	Average
TACO (1Hz unfoveated)	<i>ESF</i>	100%	100%	25%	0%	56.25%
	<i>VFH</i>	100%	100%	25%	0%	56.25%
TACO (1Hz foveated)	<i>ESF</i>	25%	100%	50%	0%	43.75
	<i>VFH</i>	75%	75%	50%	0%	50%
Kinect	<i>ESF</i>	50%	100%	100%	25%	68.75%
	<i>VFH</i>	100%	75%	75%	100%	87.5%

Table 3: Recognition results grouped by object type.

Even in such a simple recognition scenario, we can observe that the results with the TACO sensor are inferior to those reported using the Kinect sensor. Overall, foveation does not seem to provide any significant improvement, sometimes even causing the object to remain undetected compared to the linear trajectory (see Table 2). In the reported configuration, foveation increases the density of points compared to 1Hz linear trajectories by a factor of 2. Figure 69 showcases a positive recognition for both the Kinect and the TACO sensor. Unfortunately, the data obtained with 10Hz trajectories is not usable due to the “wavy surface” artifact and therefore not easy to assess improvements through foveation. The results obtained using the Kinect sensors in this scenario are good, deteriorating slightly as the distance from the target to the sensor increases due to the well-known quantization artifacts present in the sensor, especially for the ESF descriptor (see Table 2).

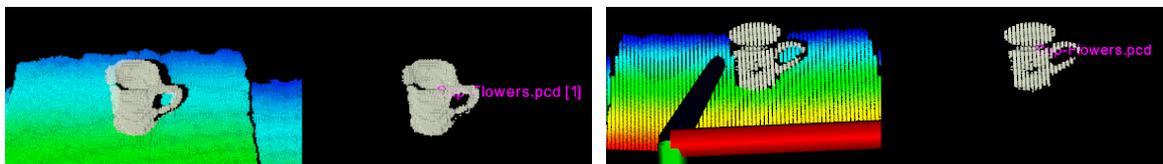


Figure 69: Positive recognition for Kinect and TACO data. Observe the difference in the point cloud reported by both sensors (Kinect: left, TACO: right)

Taking into account the reported results in such a simple scenario, it does not make sense to increase the complexity of the tests as the TACO data is clearly inferior and it is probably safe to conclude that in its current status (taking into account the aforementioned artifacts), the data it is not good enough to perform reliable object recognition. It is however interesting to mention that the quality of the TACO data depends on the material properties of the object being analyzed; for the mug and the spray bottle, the point clouds seem to be better than those reported for the tetra pack and the cylinder (see Table 3). In cases where the TACO sensor performs appropriately, it is possible to estimate the 6DOF pose of the object after correct identification. We used the OUR-CVFH [3] descriptor to showcase this in a single scene where the spray bottle is correctly recognized and its pose relative to the camera accurately estimated (see Figure 70).

Mug		1m		1.5m		2m		2.5m	
		1Hz	Fov.	1Hz	Fov.	1Hz	Fov.	1Hz	Fov.
TACO	ESF	1	0	1	0	1	1	1	0
	VFH	1	1	1	1	1	0	1	1
Kinect	ESF	1		1		0		0	
	VFH	1		1		1		1	

Spray bottle		1m		1.5m		2m		2.5m	
		1Hz	Fov.	1Hz	Fov.	1Hz	Fov.	1Hz	Fov.
TACO	ESF	1	1	1	1	1	1	1	1
	VFH	1	1	1	1	1	0	1	1
Kinect	ESF	1		1		1		1	
	VFH	1		1		1		0	

Tetra pack		1m		1.5m		2m		2.5m	
		1Hz	Fov.	1Hz	Fov.	1Hz	Fov.	1Hz	Fov.
TACO	ESF	1	1	0	0	0	1	0	0
	VFH	1	1	0	0	0	1	0	0
Kinect	ESF	1		1		1		1	
	VFH	1		0		1		1	

Cylinder		1m		1.5m		2m		2.5m	
		1Hz	Fov.	1Hz	Fov.	1Hz	Fov.	1Hz	Fov.
TACO	ESF	0	0	0	0	0	0	0	0
	VFH	0	0	0	0	0	0	0	0
Kinect	ESF	1		0		0		0	
	VFH	1		1		1		1	

Table 4: Raw recognition results for the different objects used during testing

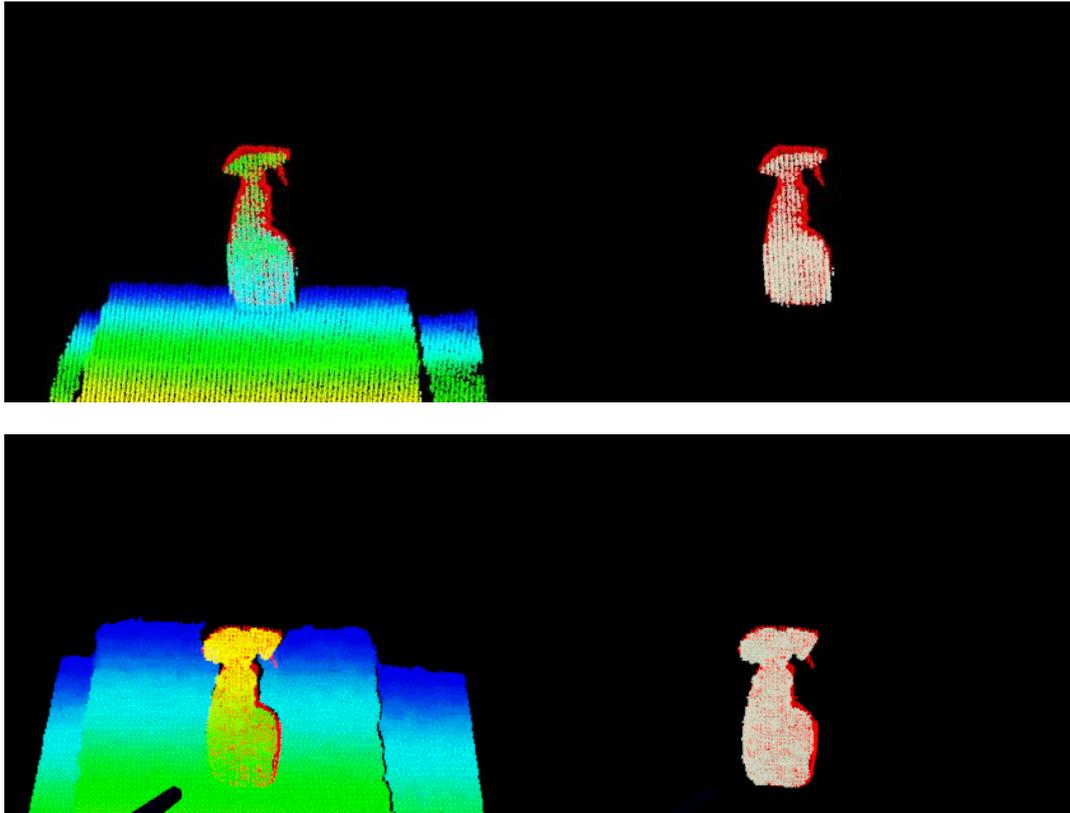


Figure 70: Correct identification and 6DOF pose estimation for the spray bottle (TACO: top, Kinect: bottom)

3.3.2.5 Effect of direct sunlight (Kinect vs TACO)

In this section, we provide a qualitative overview of the effect of direct sunlight on both the Kinect and the TACO sensor. In order to do so, we configured a static scene containing two of the objects used in the previous experiments (mug and spray bottle). The room where the experiments were done has a window that allows direct sunlight to come in from 2pm (almost no sunlight hitting the scene) to 3pm (partial sunlight illuminating the scene). During this period of time, we captured different point clouds of the static scene using the Kinect and the TACO sensor to observe the evolution of the point clouds as more and more sunlight hits the scene with the objects of interest. Figure 71 shows the evolution of the Kinect point cloud during this period. Observe how as the amount of sunlight increases, more and more parts of the scene are not reported by the Kinect sensor; eventually causing the spray bottle to completely disappear in the last point cloud (top-right in Figure 71).

Figure 72 and Figure 73 respectively show TACO (foveated) and Kinect point clouds obtained at around 2pm (left) and 3pm (right). It is clearly visible that the TACO point cloud remains unaffected under the influence of light allowing the robot to detect the two objects of interest (see Figure 74). These qualitative results suggest that the TACO sensor might be successfully used outdoors or in domestic environments with windows allowing sunlight to come in once the rest of artifacts and problems with the sensor are mitigated.

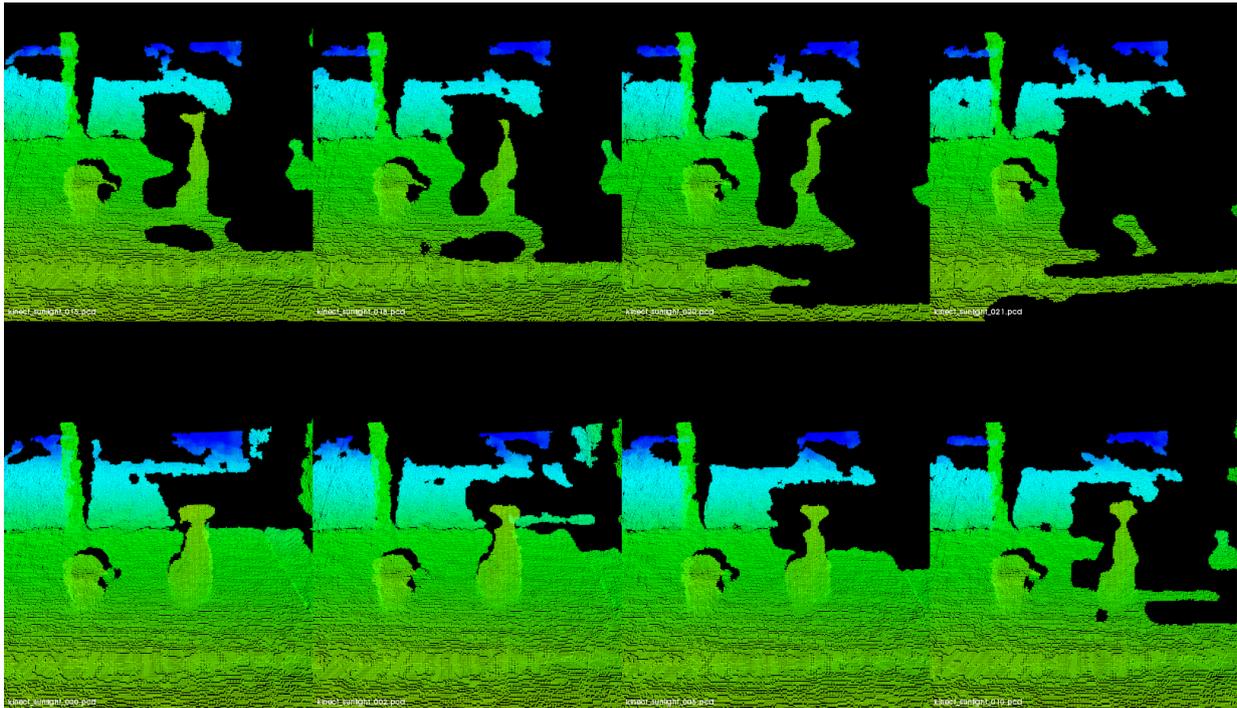


Figure 71: Evolution of the Kinect point cloud under the effect of direct sunlight. The figure shows 8 point clouds (from bottom-left to to top-right) of a static scene captured at different timestamps as the sunlight hits the scene through an opened window. Observe how the dark areas grow as more sunlight hits the field of view of the kinect.

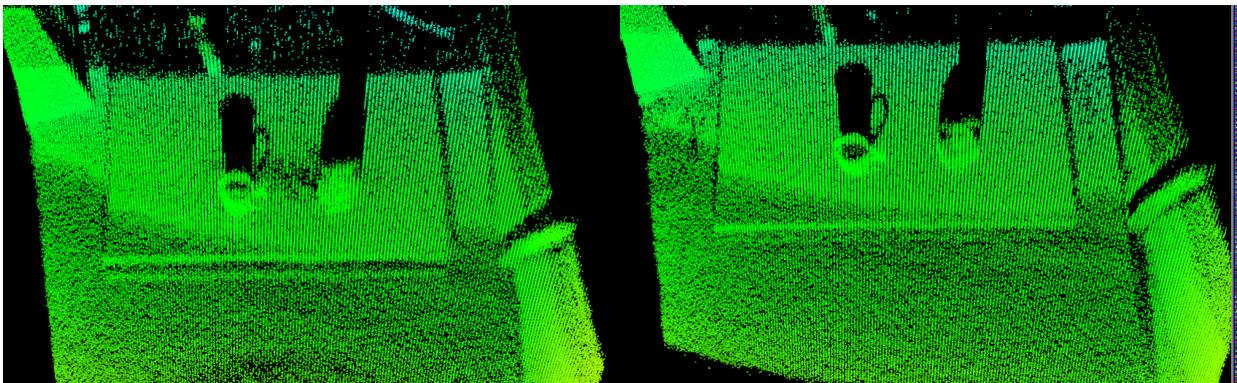


Figure 72: Point clouds obtained from the taco sensor (left: no sunlight, right: partial sunlight). The image allows seeing the foveated area with a denser resolution (spanning over the objects to be recognized). Observe how the point cloud on the right remains unaltered.

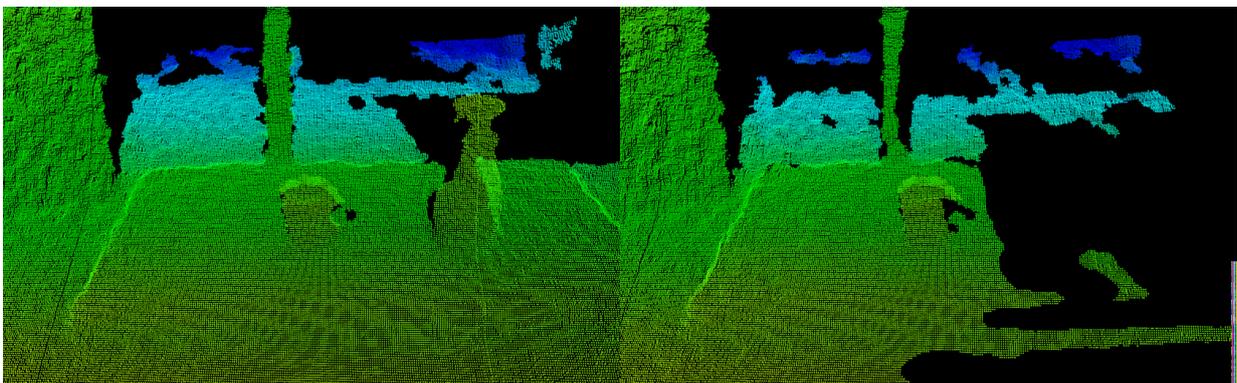


Figure 73: Point clouds obtained from the Kinect sensor (left: no sunlight, right: partial sunlight).

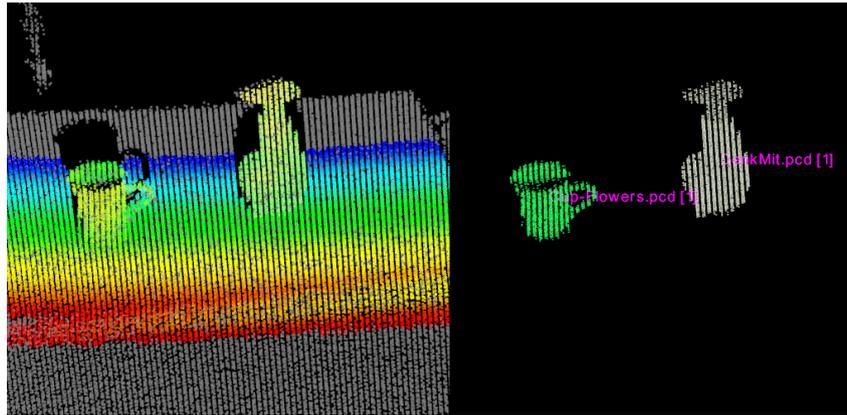


Figure 74: The mug and the spray bottle are correctly recognized using the point cloud obtained from the TACO sensor under partial sunlight. The coloured point cloud on the left part of the figure shows the point cloud used for recognition; the rest is discarded. In the right part, the segmented objects are shown with the recognition results overlaid

3.3.2.6 Discussion

The data delivered by the TACO sensor does not seem to be appropriate for reliable object recognition in its current status. Several artifacts have been observed (wavy planar surfaces at 10Hz trajectories, curved planar surfaces, shadows) that made the deployment of the use case challenging. To overcome this and in order to provide some results regarding object detection, we reduced the complexity of the use case to a minimum. Even in such a simplified scenario, only trying to retrieve the identity of the objects, the TACO sensor turned out to be inferior to the Kinect sensor for the task at hand.

As expected, the TACO point cloud remained unaltered under direct sunlight while Kinect data quickly deteriorates or is completely missing. However, in the current status, the TACO sensor cannot be used outdoors (or under direct sunlight) for the task of object detection due to its poor results under normal indoor illumination conditions due to the aforementioned artifacts.

Regarding foveation, we did not observe any significant improvement compared to 1Hz linear trajectories (foveated region sampled at doubled resolution compared to 1Hz linear) for the recognition task. The artifacts in the data seem to have a bigger impact for recognition than any increase in resolution. Unfortunately, because the 10Hz data turned out to be unusable for this scenario, we cannot assess improvement between foveated and unfoveated 10Hz data.

Nevertheless, once the artifacts are fixed (especially those regarding the wavy planar surfaces), we would expect to rapidly detect the ground or table plane with 10Hz data and dynamically foveate on the image region where the objects on the table are found, resulting in an increase of 20 times in resolution that would definitely boost recognition compared to the 10Hz data. Especially for small objects or objects with fine structures such as the handle of a mug where sensors like the Kinect have several problems (see for instance the handle in Figure 73 that is hardly visible). Another possible workflow would consist of using the Kinect sensor for coarse analysis instead of the 10Hz linear trajectory data to steer the foveation mechanisms.

4 Does Foveation Make Sense?

One unique characteristic of the TACO sensor is its ability to foveate into different parts of the scene in order to provide a denser point cloud of the region of interest. Even though the final foveation capabilities of the sensor are distant from those in the initial specification, quite an effort has been put in the development of foveation during the project. Please recall that the sensor in its current state is able to foveate in four different parts of the image (lower, semi-lower, semi-upper and upper) using 1Hz trajectories; which means that any of these four parts of the image can be analyzed approximately 20 times more densely compared to the 10Hz linear trajectories (see Figure 75).

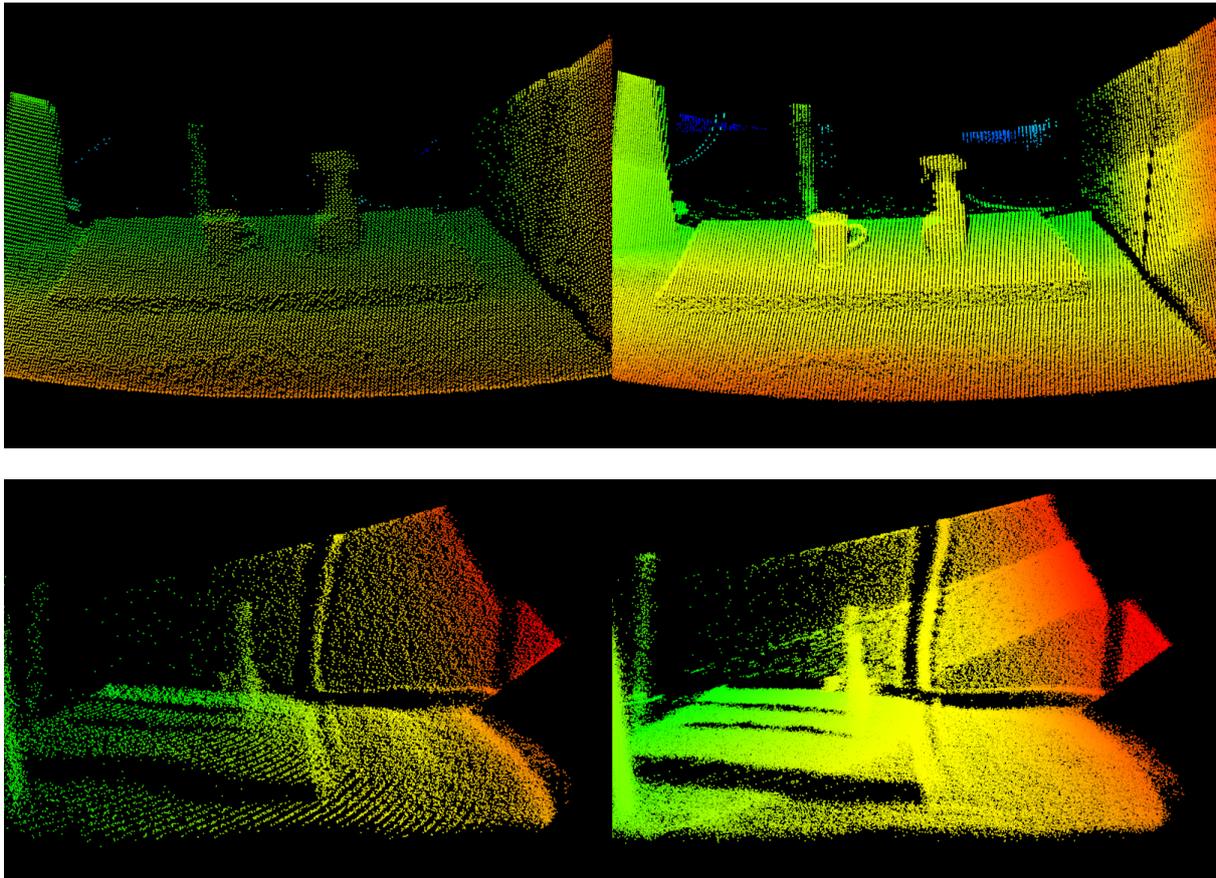


Figure 75: Comparison between unfoveated (10Hz linear trajectory) and foveated (Fov 3 - 1Hz trajectory – semi-upper part of the image). Observe the denser band of points on the foveated image. Top: front view, Bottom: side view. In the side view is it again possible to observe the wavy planar surfaces. The effect is mitigated but still visible in 1Hz trajectories.

Now that the project is coming to an end, we need to ask ourselves if foveation makes sense at all. In this section, we summarize and analyse the outcome of the different use cases related to the foveation capabilities of the sensor and try to answer the question naming this section. In section 5 "Application map", additional applications where the foveation concept might be useful are listed and further analysed.

4.1 Summary from the uses cases regarding foveation and discussion

The majority of the use cases were designed in order to show the potential of foveation for different applications. Unfortunately, several artifacts were present in the data obtained from the TACO sensor that made the deployment of foveation in practical situations very challenging. The curved planar surfaces, shadows with high laser power and in general noisy data (both intensity and range data), especially in 10Hz trajectories, did not permit to any of the end-users to actually test dynamic foveation.

Nevertheless, there is still a strong belief in foveation from the end-users partners as a powerful tool to save resources (avoiding the use of robots to physically get closer to the targets) in order to get accurate representations of the surfaces/objects of interest. Even though, no dynamic foveation was deployed, all reported use cases ended up using the foveation modes of the sensor (by manually activating them through the ROS/sensor interface) and showed significant improvement in the data compared to the 10Hz linear trajectories.

In particular, TUW has been able to perform object recognition in the high resolution data (both 1Hz trajectory and foveated 1Hz modes) obtained from the TACO sensor. Shadow has been able to use the denser data to initialize an object tracker and successfully track the object of interest in 10Hz data. Finally, OTL was able to detect the position of several vessel damaged tiles as well as an accurate reconstruction using the foveated modes of the sensor. Unfortunately, the flaws in the intensity data coming from the sensor did not allow detecting damaged parts. Such flaws on the intensity data remain visible in the foveated modes.

To summarize, there is an agreement that foveation offers a great potential for several applications but it needs to be accompanied with useful data for a successful coarser analysis. Additionally, we believe that more advanced foveation modes (X&Y resolution increase, windowed foveation, etc.) would increase even further the utility of foveation.

5 Application map and lessons learned – the TACO sensor

This chapter describes the future use or the application map of the TACO sensor system and its subtechnologies.

We present each technology and its potential use in other systems. We also give a short description of the lessons learned during the project.

5.1 Foveation and foveating systems

- **Lessons learned:** A realization is that it is the total system that needs to solve a task. This means that top down attention is more interesting than bottom up foveation, and objects are more interesting than features since it is objects we can manipulate by for example a robot. The foveation needs therefore to be application specific. A consequence of this is that the 3D vision system cannot be disconnected from the rest of the system except in very special cases.
- Foveation is a good means for reducing data bandwidth or time used to capture data of a given quality. E.g. the TACO sensor concept.
- Foveation for changing sensors used. Start capturing using a fast coarse sensor before zooming in on relevant areas using a high quality sensor. E.g. Kinect detects interesting area and use laser triangulation or laser distance measurement in this area in detail.
- Example of applications:
 - Surveillance (detect person, foveate on face for recognition)
 - Grasping (detect rough location of object, foveate to get precise position)
 - Inspection (detect rough location of object, foveate to get precise position)
 - In general: Foveating systems make sense if one has one system which is slow and precise, which can be sped up by using a fast and imprecise system. This means that the sensing principles of each system could be different.
 - In applications such as personal robotics where mobile platforms are required to search for specific objects, sensors with foveation capabilities can speed up the process when combined with attention mechanisms. The idea is to recognize from a larger distance coarse zones that are likely to contain the object of interest (planar surfaces, cupboards, etc.) to foveate on such regions without requiring the mobile platform navigating (which is a costly and slow process) to them in order to get a better resolution of the area.
 - Any type of search: search means that the target object is known and, hence, the characteristics of the target object are known. This contextual knowledge can be utilized for a foveated search procedure. The contextual information sets characteristics for the lower resolution fast first search and possible candidates are then ranked and attended to by the foveation in the next sensor cycle. Object contextual information tells about semantic information on where to find objects, e.g., pictures, power plugs, switches on walls vs. mugs, coffee machine, and plates on tables. Similarly, plates or cutlery are stored in shelves/cupboards resp. drawers. While milk is in a refrigerator again priming search for cupboards of a certain height and size. Even further, the entire layout of rooms could be modeled in this way as

humans do shown in the seminal work of M. Land [4]. A foveating sensor could be used to develop a system in this line of rapid attentional operation.

5.2 Attention algorithms

- LineMod: Detection and pose estimation of a given object
 - Visual servoing, for grasping: pick hanging fruit
 - Handle or grasping of known objects on oil platforms
 - Many objects in industry are 3D featureless
- Stereo LineMod: Use 2D LineMod to reduce false positives when detecting objects in stereo images.
 - Perform better than regular stereo in object detection
- Contour tracker: Tracks 3D featureless objects using their depth contour, for example book on a table cases
 - Visual servoing, for grasping: pick hanging fruit
 - Many objects in industry are 3D featureless
- Range model: Detect changes in range image with adaptive background learning rate
 - Surveillance. Zoom in on moving objects in the scene
- Egomotion and navigation:
 - Foveation not very relevant – mostly done due to DoW
- Planar surface segmentation and attention mechanisms:
 - Segmentation done in unfoveated data, foveate on the regions defined by the attention mechanism.
- Jump and Roof Edges:
 - Generic object to background separation. Jump edges characterize object borders while convex roof edges indicate an edge on an object. Concave roof edges may belong to a concave object or indicate in most cases an object separation.

5.3 Real-time mirror control

- **Lessons learned:** Latency is a challenge. The analysis and therefore the actual sensor's foveation is naturally lacking behind the real world. Control loop and vision system should be as close as possible on the sensor to reduce the latency of this effect.
- Are geared towards this specific TACO sensor. Reuse is not likely.

5.4 Foveation Hardware

- **Lessons learned:** The 3D sensor concept of foveation used in TACO - that is acquiring distance images at video like frame rates with coarse spatial resolution, rapidly detecting regions of interest (ROI), and then concentrating further image acquisition on these ROIs with adaptive scanning – required a challenging 2D scanning device with quasi-static actuation, large effective aperture $\geq 5\text{mm}$, and large FOV $> 60^\circ$ beyond the state of the art. The best technical compromise of the fast adaptive scanning unit were found by IPMS in a quasistatic / resonant raster scanning MEMS mirror to (i) meet opposite requirements of fast scanning ($> 1000\text{Hz}$), large optical scan range combined with large single mirror aperture and (ii) to enable partially foveation by adaptive vertical scanning. The 2D raster scanning mirror was designed up to the existing limits of physical properties and the limits of MEMS technology at IPMS. Especially the physical need of large reception aperture of $\leq 5\text{mm}$ could be technically solved only by segmentation into a MEMS array structure consisting of 5 single mirror elements resulting in large efforts for MEMS assembly, driving control for MEMS synchronization and optical system design (hybrid assembled optical waveguides) to deal with the distributed single mirror apertures and a sophisticated pulsed ToF distance measuring system
- The overall complexity of the current TACO system is large, hence the principle potential for MEMS scanner technology regarding low price and miniaturization required e.g. for mobile robots in home environments, could not be demonstrated with the current 3D-sensor. To enable a major economic impact the overall sensor system should be simple and good performing at the same time. The basic idea of an adaptive 3D camera with foveation is still attractive. This feedback was given to IPMS by sever industrial companies, but the technical effort have to be reduced to enable a broad commercialization. TACO was focused on a system at the technical limits. From the economical point of view it would be better to focus on a system were the overall advantages of the sensor system would be the main objective of development. But even the actual 3D-sensor is not a commercial product jet; all the subtechnologies could take a benefit by its own.
 - Mirror control still needs attention to come to a satisfactory state and to achieve safe array synchronization (larger feed sampling frequencies required, alternatively adapted low pass filtering in the mirror controller, working control loop for quasi-static motion). At the project's end it is still not clear if array synchronization could be achieved by TACO. Effort for distance measurement and control of mirror array out of proportion for indoor application due to small overall aperture despite array approach. Small aperture critical for outdoor applications.
 - Two-way foveation (amplitude reduction of the resonant motion plus quasi-static controlled motion) is a challenge for micro system implementation and laser safety, but probably required to make a convincing case for foveation.
 - Laser wavelength (1500 nm) not suitable for face recognition security applications due to strong absorption by water.

- Properties of laser source (narrow bandwidth) combined with small scanning mirrors lead to strong speckle effects which preclude reliable intensity measurement.
- Indoor: competition of structured light and TOF approaches (Kinect, MESA, IMF/PDM), not clear how price for a micro-mirror, pulse-TOF approach can be brought down.
- Outdoor: security applications necessary at least 100 m (black target) range to watch premises around sensitive property, similar for autonomous vehicle control, desirably more. Competition here from (still US military, but existing) InGaAs single photon detection TOF cameras which could use zoom instead of foveation (price currently 250 kEUR, TACO 60-70 kEUR). Aperture (or sensor sensitivity) must go up to achieve range, but then it is not clear if foveation can be sufficiently fast to replace multiple sensors. More research is needed.
- Outdoor: agriculture could be possible application (control of vehicles, crop watching, weed identification, etc.), but price must come down. Unclear how to replace fiber amplifier as light source. Laser diode stacks have large beam exit diameters not easily combined with scanning micro-mirror devices.

5.5 Applications for the developed hardware

- Outdoor and safety applications require larger mirror aperture, achievable with resonant mirror arrays and motor control of a 'quasi-static' axis.
- High-repetition rate electronically controlled laser pulse generation and time detection circuitry can be used in applications requiring low-jitter picosecond electronics (fast fluorescence measurements, precise localization in fibers (OTDR)).
- TOF circuitry can in principle be used in applications for infrastructure monitoring, in niche cases, where phase measurement is not sufficiently robust

5.6 Applications for the developed MEMS scanner technology

The quasistatic micro mirror technology – developed by Fraunhofer IPMS – is very promising for several industrial or consumer applications besides 3D-ToF sensing. Even though the current MEMS scanners with vertical out-of-plane comb drives are limited to 1D-quasistatic actuation or 2D-raster scanning – where partial foveation can be realized only in vertical direction by adaptive linearized scanning and horizontal direction is limited to resonant scanning only – these novel electrostatic raster scanning mirrors are a progress compared to state of the art MEMS technology. Several industrial contacts exist at IPMS due to the MEMS developments in TACO focusing on application specific MEMS developments based on the quasistatic / resonant 2D-raster scanner principle used in TACO. Electrostatic actuation in general is attractive for price sensitive applications with larger quantities due to the cost efficient fully on-chip integrated driving and sensing parts of the MEMS scanner. Compared to state of the art bi-resonant MEMS scanner the 2D-raster scanning mirror technology of IPMS enables a significant larger optical resolution with flexible frame rates due to the linear scanning in vertical direction – required by many industrial applications.

The perspective fields of application for 2D-raster scanning mirrors developed by IPMS are:

- Laser projection for (i) head up laser displays for automotive and (ii) highly miniaturized and portable laser (pico) projectors for consumer products (e.g. miniaturized laser scanning display integrated into mobile devices like smartphones),
- Medical imaging: e.g. fast confocal fluorescence microscopes or novel portable diagnostic systems based on flying spot imaging
- Miniaturized or portable laser systems (if MEMS mirror is coated with high reflective bragg-coatings) used e.g. for fast adaptive beam guidance in laser medicine for cutting of hard tissue or in industrial laser processed (e.g. in stereo lithographic systems for rapid-prototyping or miniaturized adaptive scanning system for laser marking of polymers).
- Miniaturized or faster OCT systems e.g. in ophthalmology for observation of retina diseases

- Miniaturized 3D-sensors with adaptive resolution, more suited for shorter distances (<5m) to reduce technical efforts for reception aperture and array synchronization.

5.7 Applications for alternative hardware

- There is a clear need in the security sector for an imaging sensor, either passive or in combination with illumination in a spectral region not accessible to silicon photosensors. Alternative scanning optics with larger aperture is required to address this need.
- There is still a need of robust and fast scanning mirrors with large single aperture > 5mm especially with the ability for vector scanning (two dimensional quasi-static actuation). These vector scanning units must fulfill dynamic properties of vector scanning similar to the commonly used galvanometer scanner systems. Therefore the scanning system requires at the same time a high eigenfrequency (< 1kHz) and a large static tilt angle >10°. For large apertures > 5mm required for 3D-ToF sensors, these cannot be realized with an electrostatic actuation (e.g. used in TACO) due to the limited electrostatic driving forces. First conceptual investigations for improved large aperture MEMS exists at IPMS, promising to meet the current properties of the TACO sensor with a significant reduced effort, but a new development (project) is required make these reality.

5.8 Exploitation

SINTEF has already ongoing several industrial projects using 3D vision and analysis for:

- ...simultaneous navigation and manipulation in complex dynamic scenes (SeamLess)
- ...robot's cooperation actions (NextGenRob)
- ...for robot navigation and manipulation in harsh environments (MesaVerde –Statoil's robotic platform)

TUW could apply and extend the developments in the following projects:

- ... top down language-based attention in a human robot collaboration scenario using incremental language processing (inSitu)
- ... grasping of objects from random piles (GRASP)
- ... and sorting clutter with small robots in child rooms (Squirrel, starting 2014)

5.9 General lessons learned

The project would have benefited from not developing everything in parallel. The risk is unnecessarily high when there are no integrations of data capturing hardware and data analysis software during the project period. The cause for this is of course the amount of work in getting the sensor up and running and the timeframe and format of EU projects. To reduce this risk simulated data was captured early in the project and these data are the input for all analysis software development. Deviations in these data from the one given by the TACO sensor are therefore not handled very well. Known differences in the data, like the effect of the TACO sensor's scanning mirror trajectories, are simulated and its effect is taken into account in the simulated data used. The communication protocol between the TACO sensor, the TACO foveation software and the robot (the user) was early determined and simulated. This has eased the integration of TACO hardware and software fairly easy.

6 References

- [1] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010
- [2] Walter Wohlkinger and Markus Vincze, "Ensemble of Shape Functions for 3D Object Classification", in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2011
- [3] Aitor Aldoma, Federico Tombari, Radu Bogdan Rusu and Markus Vincze, "OUR-CVFH: Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation", DAGM-OAGM, 2012
- [4] Michael Land, Neil Mennie, Jennifer Rusted: The roles of vision and eye movements in the control of activities of daily living; *Perception*, 1999, volume 28, pages 1311-1328.

7 Appendix

7.1 Public Safety Use Case Data

7.1.1 TACO Sensor, Normal Light

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371559290.34853		285	144	0	519	141	234	234	50.53%	519	163.63	0.00	7.18
1371559292.388950		258	123	0	452	135	194	194	47.67%	452	127.07	0.00	5.98
1371559294.401567		285	136	0	542	149	257	257	47.72%	542	163.72	0.00	7.03
1371559296.302275	hand in	228	126	0	591	102	363	363	55.26%	591	254.96	0.00	10.49
1371559298.591053	moving	302	147	0	2269	155	1967	1967	48.68%	2269	1,023.77	0.00	21.49
1371559300.405029	large motion fragmenting cloud? Yellow quite offset	182	59	0	1856	123	1674	1674	32.42%	1856	560.49	0.00	13.01
1371559302.287410	Denser object cloud, moving towards sensor?	510	172	0	1941	338	1431	1431	33.73%	1941	529.54	0.00	12.02
1371559304.524592		265	147	0	1383	118	1118	1118	55.47%	1383	693.93	0.00	18.66
1371559306.636209		116	43	0	1538	73	1422	1422	37.07%	1538	542.28	0.00	13.83

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371559308.790730	Moving towards and low again	504	265	0	2223	239	1719	1719	52.58%	2223	1,026.17	0.00	21.76
1371559310.402400		309	165	0	2528	144	2219	2219	53.40%	2528	1,267.64	0.00	25.21
1371559312.291141	Back on table, hand on end.	253	157	0	2491	96	2238	2238	62.06%	2491	1,482.22	0.00	29.70
1371559314.392411	No hand.	334	176	0	543	158	209	209	52.69%	543	162.95	0.00	6.99

7.1.2 TACO Sensor, Strong Light

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371559547.82168	Not moving.	335	154	0	570	181	235	235	45.97%	570	148.02	0.00	6.20
1371559549.340983		353	184	0	609	169	256	256	52.12%	609	191.21	0.00	7.75
1371559551.168856		366	185	0	593	181	227	227	50.55%	593	166.77	0.00	6.85
1371559553.121230	Hand in scene, not touching	380	201	0	1586	179	1206	1206	52.89%	1586	730.49	0.00	18.34
1371559555.306436	Hand on object and object moving	325	173	0	2699	152	2374	2374	53.23%	2699	1,350.25	0.00	25.99
1371559557.290592		154	85	0	1713	69	1559	1559	55.19%	1713	905.41	0.00	21.88
1371559559.406953		325	169	0	2515	156	2190	2190	52.00%	2515	1,220.84	0.00	24.34
1371559561.356796		194	94	0	1906	100	1712	1712	48.45%	1906	872.56	0.00	19.99
1371559563.330002		385	198	0	2511	187	2126	2126	51.43%	2511	1,186.97	0.00	23.69
1371559565.308555		349	188	0	2339	161	1990	1990	53.87%	2339	1,165.67	0.00	24.10
1371559567.432643		296	176	0	2578	120	2282	2282	59.46%	2578	1,456.29	0.00	28.68
1371559569.185697		214	130	0	2142	84	1928	1928	60.75%	2142	1,246.89	0.00	26.94

7.1.3 TACO Sensor, Darkness

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371573198.485646		234	115	0	332	119	98	98	49.15%	332	73.69	0.00	4.04
1371573200.451035		200	98	0	310	102	110	110	49.00%	310	78.82	0.00	4.48
1371573202.35504		251	125	0	361	126	110	110	49.80%	361	83.80	0.00	4.41
1371573204.285985	Hand in and motion	194	119	0	4480	75	4286	4286	61.34%	4480	2,700.78	0.00	40.35
1371573206.367330		228	144	0	3570	84	3342	3342	63.16%	3570	2,199.45	0.00	36.81
1371573208.371271		325	194	0	3258	131	2933	2933	59.69%	3258	1,861.63	0.00	32.61
1371573210.387196		452	271	0	3395	181	2943	2943	59.96%	3395	1,917.56	0.00	32.91
1371573212.382551		381	215	0	3965	166	3584	3584	56.43%	3965	2,138.42	0.00	33.96
1371573214.386113		205	122	0	3825	83	3620	3620	59.51%	3825	2,225.32	0.00	35.98
1371573216.385981		412	216	0	3682	196	3270	3270	52.43%	3682	1,821.21	0.00	30.01
1371573218.472142		190	110	0	3511	80	3321	3321	57.89%	3511	1,984.87	0.00	33.50

7.1.4 Kinect Sensor, Normal Light

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371568380.237187	Not moving	763	646	0	21750	117	20987	20987	84.67%	21750	18,312.72	0.00	124.17
1371568382.362631		749	542	0	21249	207	20500	20500	72.36%	21249	15,222.73	0.00	104.43
1371568384.590638		712	582	0	23295	130	22583	22583	81.74%	23295	18,932.71	0.00	124.05
1371568386.324016	Hand in, slight move	538	405	0	29337	133	28799	28799	75.28%	29337	21,983.02	0.00	128.35
1371568388.386073		602	409	0	29788	193	29186	29186	67.94%	29788	20,105.08	0.00	116.49
1371568390.684881	Faster motion, tracker off.	551	306	0	29074	245	28523	28523	55.54%	29074	16,008.85	0.00	93.89
1371568392.584541		419	339	0	29244	80	28825	28825	80.91%	29244	23,594.94	0.00	137.98
1371568394.379098	Middle of object cloud missing	264	235	0	27452	29	27188	27188	89.02%	27452	24,410.40	0.00	147.33
1371568397.495650		351	330	0	28536	21	28185	28185	94.02%	28536	26,808.74	0.00	158.70
1371568398.965245		440	316	0	30224	124	29784	29784	71.82%	30224	21,616.33	0.00	124.34
1371568400.114221		474	358	0	30366	116	29892	29892	75.53%	30366	22,846.00	0.00	131.10
1371568402.350045		508	332	0	30895	176	30387	30387	65.35%	30895	20,074.95	0.00	114.21

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371568404.351993		460	398	0	31385	62	30925	30925	86.52%	31385	27,100.52	0.00	152.97
1371568406.362834	Back on desk, with hand	590	422	0	31853	168	31263	31263	71.53%	31853	22,661.22	0.00	126.97
1371568408.830431		552	484	0	32037	68	31485	31485	87.68%	32037	28,029.88	0.00	156.60
1371568410.369797	Not moving	724	537	0	27996	187	27272	27272	74.17%	27996	20,623.58	0.00	123.26

7.1.5 Kinect Sensor, Strong Light

No usable result.

7.1.6 Kinect Sensor, Darkness

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371571538.072257		735	605	0	22952	130	22217	22217	82.31%	22952	18,782.56	0.00	123.98
1371571542.332182		783	623	0	22533	160	21750	21750	79.57%	22533	17,797.63	0.00	118.56
1371571544.303320		756	579	0	22621	177	21865	21865	76.59%	22621	17,185.69	0.00	114.26
1371571546.390413		750	607	0	22960	143	22210	22210	80.93%	22960	18,463.42	0.00	121.85
1371571548.356247	Hand in. Object moving	606	475	0	31935	131	31329	31329	78.38%	31935	24,927.33	0.00	139.49
1371571550.403640	Fast motion	552	400	0	33215	152	32663	32663	72.46%	33215	23,957.35	0.00	131.45
1371571552.303064		565	455	0	32280	110	31715	31715	80.53%	32280	25,905.55	0.00	144.19
1371571554.324746		529	473	0	31857	56	31328	31328	89.41%	31857	28,433.79	0.00	159.31
1371571556.580496		438	366	0	30829	72	30391	30391	83.56%	30829	25,700.33	0.00	146.37
1371571558.297578	Missing middle of object cloud	288	231	0	30664	57	30376	30376	80.21%	30664	24,549.02	0.00	140.19

bag_time	notes	true_points	tp_points	fp_points	target_points	fn_points	false_points	tn_points	accuracy	n	chi_squared	p_value	phi
1371571560.320186		345	288	0	30750	57	30405	30405	83.48%	30750	25,621.53	0.00	146.11
1371571562.479610		537	412	0	31917	125	31380	31380	76.72%	31917	24,390.37	0.00	136.52
1371571566.289353		507	443	0	32034	64	31527	31527	87.38%	32034	27,933.56	0.00	156.07