

# **521479S Software project (2025) - Hack the dataset from X200GO - 3D modelling sensor**

## **Hack the dataset from X200GO -3D modelling sensor Final Report**

1.0 Version

Name	Student number
Sami Häkkinen	
Syed Amin Al Mahim	
Muhammad Huzaifa	

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

## 1. Executive Abstract

This work addresses the challenge of extracting structured information from undocumented binary file formats originating from a Simultaneous Localization and Mapping (SLAM) system. These files encapsulate sensor data, including inertial measurements, raster imagery, global navigation satellite system (GNSS) positioning, and LiDAR-based spatial information (point clouds). Understanding of such raw data can be useful for enabling advanced computational approaches, such as algorithmic optimization and machine learning, which can enhance mapping accuracy and visualization beyond the capabilities of proprietary software solutions.

The lack of manufacturer-provided specifications for these proprietary file formats necessitates a systematic methodology for reverse engineering. This project explores established techniques for analyzing unknown binary structures, drawing on insights from scientific literature, and prior research in data parsing and sensor fusion [7]. In addition, it presents our practical strategies developed during the project, detailing the tools, technologies, and custom software implementations employed to interpret and extract meaningful content from the data. By providing a framework for understanding and processing these data sources, the work aims to facilitate innovation in SLAM-related applications and support the development of more robust and adaptive algorithms.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

## Contents

<b>1. EXECUTIVE ABSTRACT</b>	<b>2</b>
<b>2. INTRODUCTION</b>	<b>4</b>
<b>3. CUSTOMER</b>	<b>5</b>
<b>4. ACRONYMS AND DEFINITIONS</b>	<b>5</b>
<b>5. PROJECT DESCRIPTION</b>	<b>5</b>
5.1 Approaches	6
5.1.1 Binary Inspection	6
5.1.2 Frame Based Testing	8
5.1.3 GNSS & NMEA Pattern Recognition	9
5.1.4 Visualisation Based Validation	11
5.1.5 Reverse Engineering	11
5.2 Discovery of White Label Application	12
5.3 Results	12
5.4 Deployment	13
5.5 Challenges	13
5.5.1 Lack of Documentation	13
5.5.2 Proprietary and Potentially Obfuscated Data	13
5.5.3 Validation Without Ground Truth	13
5.5.4 High Computational Complexity	13
<b>6. REALISED REQUIREMENTS</b>	<b>13</b>
6.1 Binary Data Decoding	14
6.2 Development of Experimental User Interface	14
6.3 Documentation & Knowledge Transfer	14
<b>7. USED SOFTWARE DEVELOPMENT METHOD</b>	<b>15</b>
<b>8. SOFTWARE ACCEPTANCE</b>	<b>15</b>
<b>9. PROJECT TIMELINE</b>	<b>16</b>
<b>10. FEEDBACK</b>	<b>17</b>
<b>11. REFERENCES</b>	<b>18</b>

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

## 2. Introduction

Advancements in mobile mapping and autonomous sensing technologies have led to the widespread use of multi sensor platforms that combine LiDAR, Global Navigation Satellite Systems (GNSS), Inertial Measurement Units (IMU) and camera sensors to generate accurate three dimensional (3D) representations of environments [3]. One such system is the Stonex X200 GO (*Fig: 1*), a portable SLAM based 3D mapping sensor designed for rapid data acquisition in both indoor and outdoor environments [8].

The X200 GO system integrates multiple sensors, including a rotating LiDAR scanner, cameras, IMU and GNSS. During operation, the sensor records large volumes of raw data that are later processed using the manufacturer's proprietary software, *Slam Go Post*, to produce point clouds, trajectories and raster images [8]. However, the raw sensor outputs are stored in proprietary and undocumented binary file formats, which prevents direct access to the underlying data without relying on the vendor's closed source software.



*Figure 1: Stonex X200 GO*

The primary objective of this project is to reverse engineer the binary data formats generated by the X200 GO sensor [8] and to develop software tools capable of parsing and converting these files into human readable and standardized formats. Achieving this would enable independent data analysis, visualization and integration with open source tools, while also supporting further research and algorithm development beyond the constraints of the proprietary system.

Unlike conventional software development projects with predefined specifications, this work is inherently research-oriented and exploratory in nature. No official documentation or reference implementation was available for decoding the X200 GO binary files, and there was no clear or straightforward method to interpret the stored sensor data. As a result, the project required a brute-force and hypothesis-driven approach, where multiple decoding strategies were formulated, tested, and iteratively refined based on observed patterns and validation results.

The X200 GO generates multiple file types, each corresponding to different sensor modalities, including LiDAR point cloud data, IMU measurements, GNSS navigation data, and raster image data. These files typically consist of a binary header followed by structured data records, often combining raw binary payloads with embedded ASCII metadata. Understanding the internal structure of these files therefore required systematic experimentation rather than direct implementation.

To address this challenge, the project adopts an experimental, data-driven reverse engineering methodology, combining low level binary inspection, pattern recognition, exhaustive brute force decoding and visualization based validation. Python was selected as the primary development language due to its strong support for binary data handling, numerical analysis,

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

and data visualization. Interactive Jupyter and Google Colab notebooks were used for rapid experimentation and hypothesis testing, while standalone Python scripts were developed to ensure reproducibility and structured analysis.

While complete decoding of all sensor modalities is a complex and high risk task, the project demonstrates meaningful progress in identifying file structures, frame boundaries and plausible data representations, particularly for GNSS related navigation data [3]. The outcomes of this research-oriented effort provide a solid foundation for future work aimed at fully open and vendor independent parsing of X200 GO sensor data.

### 3. Customer

The customer for this project is Rajesh Raveendran (*Rajesh.Raveendran@oulu.fi*), affiliated with the University of Oulu. The customer's primary requirement is the ability to access and interpret raw sensor data produced by the Stonex X200 GO SLAM system without relying on the manufacturer's proprietary processing software (*Slam Go Post*). This includes extracting and understanding data from multiple sensor modalities such as LiDAR, GNSS, IMU, and camera systems, and converting the decoded outputs into human-readable or standard formats suitable for further analysis, visualization, and research.

### 4. Acronyms and Definitions

Acronym/Term	Definition
ASCII	American Standard Code for Information Interchange
SLAM	Simultaneous Localization and Mapping
LIDAR	Light Detection and Ranging
IMU	Inertial Measurement Unit
GNGGA	NMEA Standard, contains time, date, position, fix quality, number of satellites and horizontal dilution of precision.
GPRMC	NMEA Standard, contains time, date, position, positioning mode, course over ground, and speed.
GNSS	Global Navigation Satellite System
HEX	Hexadecimal
NMEA	National Marine Electronics Association

### 5. Project Description

This project focuses on the reverse engineering of proprietary binary data formats generated by the Stonex X200 GO SLAM sensor. Due to the absence of official documentation and the closed nature of the manufacturer's software, the project required research oriented and exploratory approach rather than a conventional implementation driven workflow.

The core task of the project was to analyze raw binary files produced by the X200 GO system and to develop software tools capable of extracting meaningful sensor data without relying on

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

the proprietary Slam Go Post application. The work involved systematic experimentation, pattern recognition, brute force decoding, and validation through visualization.

The project addressed multiple sensor data types, including IMU, GNSS, LiDAR, and raster image data, each of which posed different technical challenges due to variations in structure, encoding, and data representation.

## 5.1 Approaches

Given the lack of a predefined decoding strategy, multiple complementary approaches were adopted [7].

### 5.1.1 Binary Inspection

The initial inspection of the IMU file was performed using Microsoft's Hex Editor extension in Visual Studio Code. Figure (Fig: 2), below displays the first rows of the binary IMU file in hexadecimal format. On the left, we see the hex values, with each row containing 32 addresses, while on the right, the corresponding ASCII decoded text per each address value is shown. At first glance, we can identify human readable strings such as "feimarobotics-slam imu" "feima-1200-imu," "fm-i2000," as well as a datetime string that likely represents the exact start time of the SLAM project. Using Google or AI tools such as "ChatGPT-5" did not provide any reliable or useful data sources when we tried to investigate these strings found in the first row of the binary file.

00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F 10 11 12 13 14 15 16 17 18 19 1A 1B 1C 1D 1E 1F	Decoded Text
00000000 66 65 69 6D 61 72 6F 62 6F 74 69 63 73 2D 73 6C 61 6D 2D 69 6D 75 00 00 00 00 00 00 00 00 00	feimarobotics-slam-imu.....
00000020 00 66 65 69 6D 61 2D 69 32 30 30 2D 69 6D	.....feima-i2000-im
00000040 75 00	u.....
00000060 00	.....
00000080 00 46 4D 2D 49 32 30 30 00 00	.....FM-I2000..
000000A0 00	.....
000000C0 00 00 00 00 00 00 00 00 47 75 00 00 01 00 00 05 00 00 51 00 00 00 00 00 00 00 40 7F 40	.....Gu.....Q.....@..@
000000E0 00 00 00 00 00 00 00 00 00 00 00 00 C0 41 40 00 00 00 00 40 EF 40 00 00 00 20 4F 04 44 3F	.....A@.....@.....0.D?
00000100 00 00 00 40 82 15 24 3F 00	.....\$?.....
00000120 00	.....
00000140 00	.....
00000160 00	.....
00000180 00	.....
000001A0 00	.....
000001C0 00	.....
000001E0 00	.....
00000200 00	.....
00000220 00 53 75 6E 20 53 65 70 20 20 37 20 31	.....Sun Sep 7 1
00000240 35 3A 30 38 3A 32 37 20 32 30 32 35 0A 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00	5:08:27 2025.....
00000260 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 58 32 30 30 47 4F 32 35 30 35 31 35 30 00 00 00	.....X200G02505150...
00000280 00	.....
000002AB 00	.....

Figure 2: IMU data in hexadecimal

We continued examining the data file and as we scrolled further, some visible patterns began to emerge. As shown in Figure x, a distinct pattern in the ASCII representation starts at address '0x400', which corresponds to '1024' in decimal. Based on this observation, we can reasonably assume that the header of the binary file is 1024 bytes long (1 kilobyte) and that the actual data begins immediately after this header.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22



Figure 3

A closer look at figure (Fig: 3) reveals that certain characters appear at regular intervals. For example, the ‘%’ symbol occurs consistently throughout the data. In this figure, each row is 32 bytes long, which is the maximum display width for the hex editor used in the visualization. Observing the placement of the ‘%’ characters suggests that each record in the data might be 33 bytes in length. Based on this assumption, we decided to reorganize all data following the header into chunks of 33 bytes. For this purpose, we created a Python program that reads the data file, skips first 1024 bytes and then puts each 33-byte chunk into Python array.

Figure (Fig: 4) illustrates printing first 25 elements of the array in the Python program. Based on the assumption of 33 bytes long records, we started implementing the first hacking approach which is documented in the next chapter.

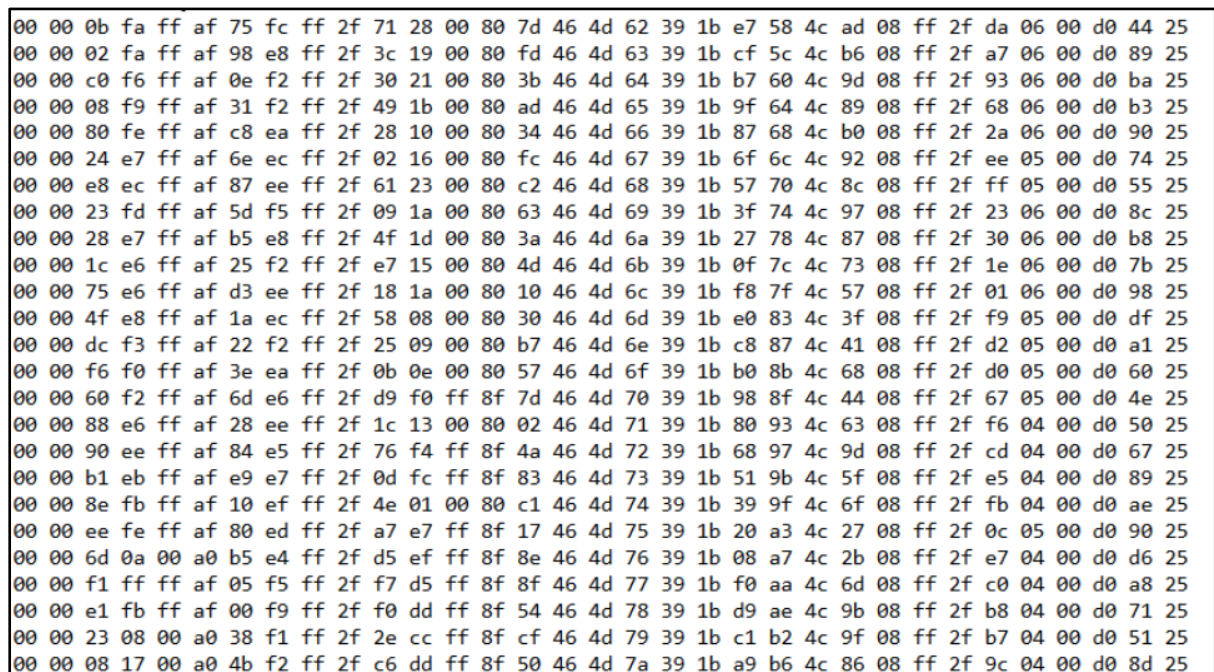


Figure 4: First 25 rows



Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

### 5.1.2 Frame Based Testing

Once the IMU data had been segmented into fixed size records, statistical analysis of the individual records became possible. As an initial approach, a visualization-driven method was adopted, where decoded values were plotted to identify plausible sensors like behavior.

Based on earlier observations, the file was assumed to contain a 1024 byte header, followed by data records of 33 bytes each. Within each record, byte sequences of two to four bytes of length were interpreted using multiple candidate data types to evaluate potential representations of sensor measurements. Five possible datatypes are:

- Signed 16-bit integers (int16)
- Unsigned 16-bit integers (uint16)
- Signed 32-bit integers (int32)
- Unsigned 32-bit integers (uint32)
- 32-bit floating point (float32)

For each data type, both little endian and big-endian formats were tested. Additionally, we vary the starting offset within the 33 byte block: for each data type, we attempt unpacking from every valid starting byte. This results in a total of:

$$5 \text{ data types} \times 2 \text{ endianness} \times 32 \text{ offsets} = 320 \text{ combinations}$$

For every possible combination, first 500 values are plotted. For example, figure (Fig: 5) shows the first 500 rows with the 20th and 21st bytes interpreted as a 16 bit signed integer in little-endian format. Figure (Fig: 6) represents 27th, 28th, 29th and 30th bytes as unsigned 32-bit integer in big-endian format. Figure (Fig: 8) shows data in Excel format where each row corresponds expected 33 bytes long record of the data and each column corresponds an offset to the first byte of the record.

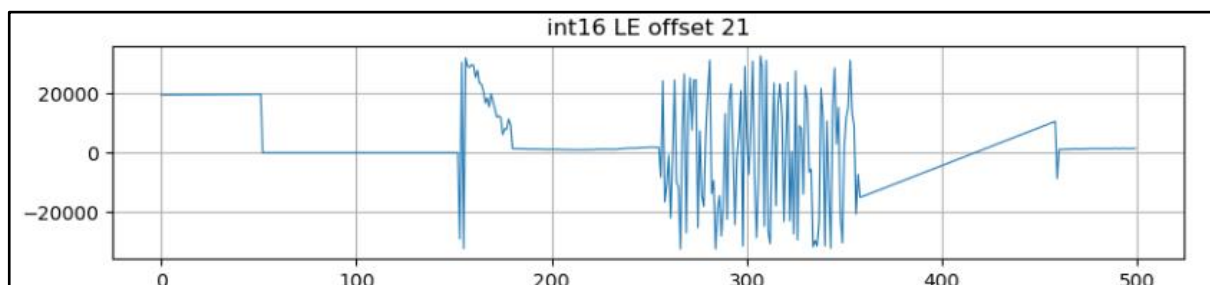


Figure 5: 20th & 21st byte interpreted for first 500 rows (16-bit signed int & little-endian order)

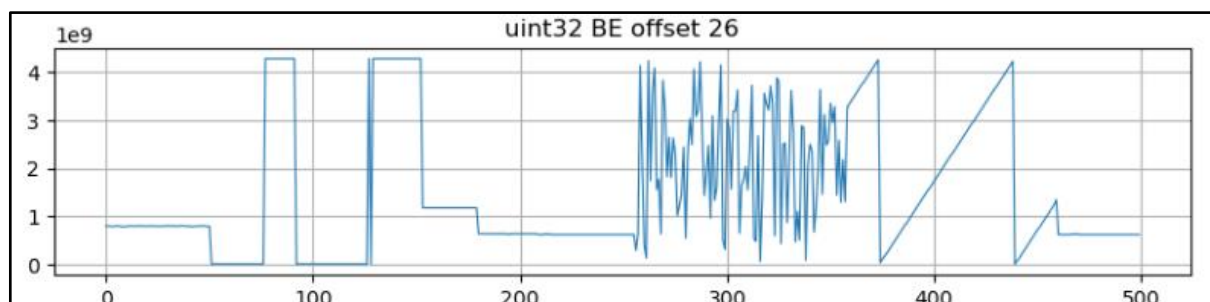


Figure 6: 27th to 30th byte interpreted for first 500 rows (32-bit unsigned int & big-endian order)





Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

	timestamp	lat_i	lon_i	alt_i	v1	v2	v3	quality		timestamp	lat_i	lon_i	alt_i	v1	v2	v3	quality	lat_deg	lon_deg	alt_m	
0	228492260264288	55065	1179649		1	0.0	0.0	0.0	0	0	228492260264288	55065	1179649	1	0.0	0.0	0.0	0	0.005507	0.117965	1.0
1	228492260264288	55066	1179648		1	0.0	0.0	0.0	0	1	228492260264288	55066	1179648	1	0.0	0.0	0.0	0	0.005507	0.117965	1.0
2	228707008629088	55115	1179649		1	0.0	0.0	0.0	0	2	228707008629088	55115	1179649	1	0.0	0.0	0.0	0	0.005511	0.117965	1.0
3	228707008629088	55116	1179649		1	0.0	0.0	0.0	0	3	228707008629088	55116	1179649	1	0.0	0.0	0.0	0	0.005512	0.117965	1.0
4	228921756993888	55165	1179650		1	0.0	0.0	0.0	0	4	228921756993888	55165	1179650	1	0.0	0.0	0.0	0	0.005516	0.117965	1.0
5	228921756993888	55166	1179649		1	0.0	0.0	0.0	0	5	228921756993888	55166	1179649	1	0.0	0.0	0.0	0	0.005517	0.117965	1.0
6	229136505358688	55215	1179650		1	0.0	0.0	0.0	0	6	229136505358688	55215	1179650	1	0.0	0.0	0.0	0	0.005522	0.117965	1.0
7	229136505358688	55216	1179649		1	0.0	0.0	0.0	0	7	229136505358688	55216	1179649	1	0.0	0.0	0.0	0	0.005522	0.117965	1.0
8	229351253723488	55265	1179650		1	0.0	0.0	0.0	0	8	229351253723488	55265	1179650	1	0.0	0.0	0.0	0	0.005527	0.117965	1.0
9	229351253723488	55266	1179649		1	0.0	0.0	0.0	0	9	229351253723488	55266	1179649	1	0.0	0.0	0.0	0	0.005527	0.117965	1.0

Figure 9: Decoded timestamps, latitude and longitude

As part of this analysis, timestamps, latitude, and longitude values were extracted from the decoded GNSS data. Multiple timestamp formats were considered during interpretation, including Unix time and other commonly used GNSS related time representations. However, the decoded timestamp values were found to be unrealistic when compared to the known data acquisition period, as the original dataset was recorded in 2025, while the interpreted timestamps did not correspond to this timeframe. This suggests that the timestamp field may use a different reference epoch, scaling factor, or encoding scheme than initially assumed.

Similarly, decoded latitude and longitude values were plotted to assess spatial plausibility. Although the values exhibited continuity and structural characteristics consistent with positional data, their absolute locations were significantly offset from the known data collection site. The dataset was recorded in Oulu, Finland, yet the plotted coordinates appeared in regions corresponding to southern Africa (*Fig: 10*).

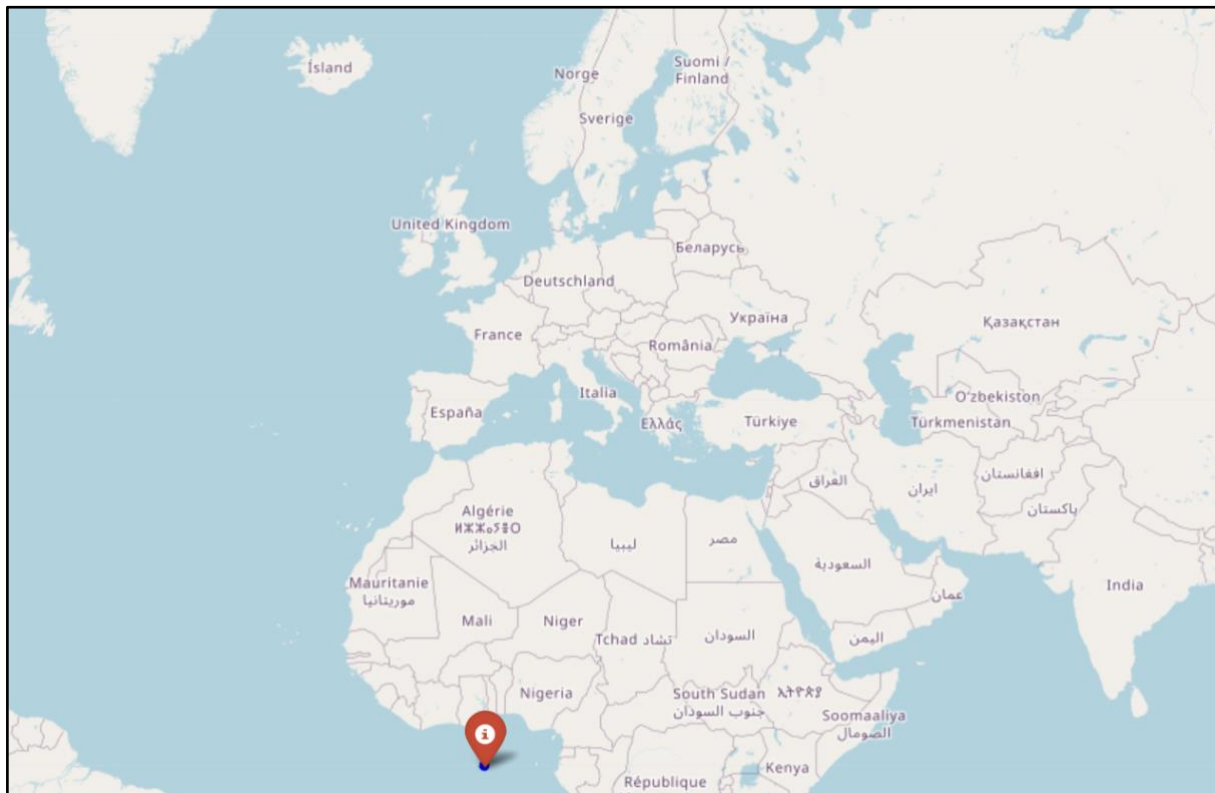


Figure 10: Decoded data when mapped.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

The GNSS and NMEA pattern recognition approach confirmed that the X200 GO navigation files contain embedded standardized GNSS messages, specifically \$GNGGA and \$GNRMC. While timestamps, latitude, and longitude fields were identifiable and structurally consistent, their decoded values could not be directly mapped to real world time or geographic location. The presence of standardized NMEA messages provides a reliable reference for synchronizing and validating binary navigation records, although full geodetic interpretation requires further investigation.

#### **5.1.4 Visualisation Based Validation**

During the initial stages of the project, an exploratory decoding approach was implemented based on direct visualization of decoded binary values. In this approach, the binary data was segmented into fixed size records and interpreted using various candidate data types and byte offsets. The resulting values were plotted as time series signals to identify patterns that might resemble sensor outputs, such as accelerometers or gyroscope measurements.

At this stage, it was assumed that meaningful sensor behavior could be detected through visual inspection of individual value streams, including characteristics such as periodic motion, noise patterns, or saturation effects. This approach was supported by visualization scripts that generated plots for multiple decoding configurations, enabling rapid comparison across different offsets and data type interpretations.

Although this method produced several visually structured plots, the results were not conclusive. Many decoded signals exhibited continuity and variation but could not be reliably associated with physical sensor quantities. Subsequent discussion with the customer revealed that this approach was inherently limited, as the X200 GO sensor operates within a three-dimensional spatial context, where meaningful interpretation depends on multi-axis correlation, coordinate frame alignment, and sensor fusion rather than isolated one-dimensional signals.

Consequently, interpreting individual decoded values without accounting for their spatial relationships and orientation in 3D space proved insufficient. The approach was therefore deemed unsuitable for further decoding of IMU and motion-related data, and the project's focus shifted toward structure-driven decoding, GNSS anchoring, and frame-level analysis.

#### **5.1.5 Reverse Engineering**

##### **5.1.5.1 Tupni**

One approach for automatic input format inference is Tupni, which uses dynamic analysis and symbolic execution to reconstruct file structures from program behavior [2]. It uses techniques such as symbolic execution, dynamic analysis, and differential fuzzing to infer the structure, constraints, and semantics of the underlying format. The result is a formal specification that can be used to build parsers or generate valid test cases. Tupni learns how the program works that handles the binary files and doesn't take the actual binary file as input.

##### **5.1.5.2 Binwalk**

Binwalk is a popular open source tool for analyzing and reverse engineering binary files [1]. It is primarily used to identify embedded file signatures, compressed data, and firmware structures within binary blobs. Binwalk can automatically scan a file for

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

known patterns, extract data sections, and provide insights into the file's internal structure. It is commonly used in embedded systems of security research, firmware analysis, and binary forensics. Binwalk attempted to analyze the IMU binary data to detect any recognizable structures, headers, or embedded sections. However, in this case, the Python environment could not import binwalk.core, preventing direct programmatic use of Binwalk. As a result, alternative approaches were pursued to parse and visualize the binary IMU frames manually.

## 5.2 Discovery of White Label Application

During the research phase of the project, an important additional finding emerged outside the original project scope. While surveying existing tools and software related to SLAM systems, the team identified an alternative application capable of decoding and visualizing data generated by the X200 GO sensor [5].

Further investigation revealed that this application is a white label variant of an existing proprietary solution, introduced under a different name and supplemented with additional features and interface modifications. Although the core functionality appeared to be derived from the same underlying technology as the manufacturer's official software, the existence of this white label version was not known to the customer prior to this project.

From a research perspective, the white label application [5] served as a reference point and feasibility indicator, reinforcing the assumption that vendor independent access to the data is achievable [5]. The discovery added tangible value for the customer by increasing awareness of alternative solutions and highlighting the broader ecosystem surrounding the X200 GO platform.

It is important to note that the existence of a white-label solution does not diminish the need for independent and open parsers. On the contrary, it emphasizes the importance of transparent, well documented decoding tools that enable reproducibility, long term maintainability, and research oriented access to sensor data.

## 5.3 Results

Despite the complexity and proprietary nature of the data formats, the project achieved several meaningful results:

- Identification of consistent binary header structures and metadata regions.
- Confirmation of fixed length frame structures for multiple file types.
- Successful partial decoding of GNSS data, including:
  - Timestamps
  - Position related fields
  - Velocity components
  - Quality indicators
- Validation of GNSS update rates and data continuity through visualization.
- Successful parsing and visualization of raster/image data using standard geospatial libraries.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

While full decoding of IMU and LiDAR payloads was not achieved, several candidate byte regions and structural patterns were identified, providing a clear direction for future work.

## 5.4 Deployment

The development was primarily in Python, leveraging its strong support for binary data handling, numerical computation, and visualization. The project utilized two main execution environments:

- Google Colab / Jupyter Notebooks: Used for exploratory analysis, rapid hypothesis testing, and visualization-based validation.
- Standalone Python Scripts: Used for structured decoding, reproducibility and batch processing of binary files.

The modular design of the scripts allows future extensions to additional file types and integration into larger processing pipelines. No external deployment environment was required beyond a standard Python runtime.

## 5.5 Challenges

Several significant challenges were encountered during the project:

### 5.5.1 Lack of Documentation

No official documentation or reference implementations were available for the X200 GO binary formats, making the decoding process inherently uncertain and experimental.

### 5.5.2 Proprietary and Potentially Obfuscated Data

Some data fields appeared intentionally encoded in non-obvious ways, complicating direct interpretation and increasing the risk of false assumptions.

### 5.5.3 Validation Without Ground Truth

Without access to raw reference values or internal sensor specifications, validation relied heavily on indirect methods such as visualization and numerical plausibility.

### 5.5.4 High Computational Complexity

Brute force decoding across multiple data types; offsets and endianness combinations required careful management to remain computationally feasible.

## 6. Realised requirements

The requirements of this project were defined with a clear distinction between primary research objectives and secondary implementation objectives. The primary objective of the project was the reverse engineering and decoding of proprietary binary data generated by the X200 GO SLAM sensor. Development of a user interface for data visualization and interaction was defined as a secondary objective, intended to be pursued only if sufficient progress was achieved in decoding the underlying data formats.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

## **6.1 Binary Data Decoding**

### **6.1.1 Requirement**

Develop binary file parsers capable of decoding and interpreting proprietary data generated by the X200 GO SLAM sensor.

### **6.1.2 Realisation**

This requirement was partially but meaningfully realized. The project successfully:

- Analyze binary file structures and headers.
- Identify frame boundaries and recurring patterns.
- Implemented brute force decoding strategies.
- Achieved partial decoding of GNSS/navigation data.

Although full decoding of all sensor modalities was not completed, the achieved results demonstrate the feasibility of vendor independent access to X200 GO data and fulfill the primary research objective of the project.

## **6.2 Development of Experimental User Interface**

### **6.2.1 Requirement**

Develop a user interface to visualize decoded data and provide basic interaction with the parsed sensor outputs.

### **6.2.2 Realisation**

This requirement was not realized during the project.

The development of a user interface was defined as a secondary objective, dependent on sufficient progress being achieved in the primary objective of decoding the proprietary binary data formats. Due to the research oriented and exploratory nature of the project, a significant portion of the available effort was required for reverse engineering tasks, including binary inspection, brute force decoding, and validation of sensor data.

As the decoding process involved high uncertainty and required iterative experimentation, prioritizing user interface development would have reduced the effectiveness of the primary research objectives. Consequently, the project team intentionally focused on understanding data structures and validating decoding feasibility rather than implementing a standalone graphical interface.

## **6.3 Documentation & Knowledge Transfer**

### **6.3.1 Requirement**

Document all approaches, findings, tools, and limitations to ensure reproducibility and future development.

### **6.3.2 Realisation**

This requirement was fully realized [6]White Label Application. Comprehensive

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

documentation was produced covering:

- Reverse engineering methodology.
- Experimental approaches.
- Achieved results.
- Discovered limitations and open challenges.

## 7. Used Software Development Method

The software development method used in this project can be best described as a research oriented, iterative, and exploratory development approach, rather than a traditional linear software engineering model. This choice was driven by the nature of the problem domain; reverse engineering undocumented and proprietary binary file formats, for which no predefined specifications, reference implementations or guaranteed solution paths were available.

At the outset of the project, it was not possible to define detailed functional requirements or a fixed implementation plan. Instead, development progressed through iterative experimentation, where hypotheses about data structures and encoding schemes were continuously formulated, tested, and evaluated. This approach aligns closely with experimental software engineering and research based prototyping practices.

Python was selected as the primary development language due to its strong ecosystem for binary data handling, numerical computation, and visualization. Jupyter and Google Colab notebooks were used extensively during the exploratory phases to enable rapid experimentation and interactive analysis, while standalone Python scripts were used to consolidate stable findings and improve reproducibility.

While elements of traditional methodologies such as Agile can be identified particularly in terms of incremental progress and continuous feedback, the overall approach is more accurately characterized as research driven iterative development, where learning and discovery were the primary drivers of progress rather than feature completion.

## 8. Software Acceptance

In this project, software acceptance was judged by how well the research goals were met, rather than by traditional user acceptance testing. The main aim was to see whether undocumented binary data formats from the Stonex X200 GO SLAM sensor could be reverse engineered, so the “acceptance tests” focused on analytical correctness, consistency of decoded data, and how reliably the same results could be reproduced.

Each decoding idea was checked against what could actually be seen in the data. The binary parsing logic was tested by confirming that file headers, frame boundaries, and recurring patterns were detected correctly across different datasets. For GNSS-related information, the decoded values were compared against known properties of NMEA messages, including their structure, field order, update rate, and continuity over time.

The implementation itself was evaluated through repeated experimentation and visual inspection. Decoded values were plotted to see if they behaved in a physically reasonable way, such as changing smoothly over time, showing realistic sampling rates, and maintaining consistent relationships between fields. Whenever results looked implausible or inconsistent, the underlying assumptions were refined or rejected so that accepted outputs were always



Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

grounded in evidence rather than chance.

Several tools supported this acceptance process [6]. Python served as the main language, offering robust libraries for binary data handling, numerical work, and visualization. Jupyter and Google Colab notebooks were used for rapid, interactive exploration, while standalone Python scripts confirmed that stable decoding approaches were reproducible. Visualization libraries helped validate GNSS trajectories and data continuity, and hex viewers were used to verify low-level interpretations of the binary files.

Even though a polished end-user application was not produced, the acceptance criteria were met by showing clear progress toward vendor-independent decoding, achieving partial but consistent GNSS data extraction, and delivering a documented methodology that others can reproduce and build on

## 9. Project Timeline

The project was planned and executed over the period from September to December 2025. Due to the exploratory nature of reverse engineering undocumented binary data formats, the project planning emphasized iterative progress, continuous validation and regular communication with the customer rather than strict feature based milestones.

The timeline was guided by an initial Gantt chart, which divided the project into research, development, documentation, and testing activities. These phases were not strictly sequential; instead, several activities overlapped to allow findings from experimentation to directly inform subsequent analysis and documentation.

The project began with familiarization phase, during which the team studied the X200 GO sensor, its role in SLAM applications and the overall decoding challenge. This was followed by a study and research phase involving a literature review on binary reverse engineering, investigation of similar SLAM systems and assessment of tools for low-level binary analysis. Initial inspection of the X200 GO data files was performed to identify headers, metadata and recurring structural patterns. The main phase of the project consisted of development and experimentation, focusing on exploratory binary analysis and brute force decoding. Multiple hypotheses regarding frame structure, data types and byte alignment were tested, with visualization-based validation.

Milestone description	Category	Assigned to	Start	Days
<b>Study &amp; Research</b>				
Project Understanding	On Track	All	9/25/2025	10
Feasibility Study	Milestone	All	10/5/2025	15
Research	Low Risk	All	10/20/2025	10
<b>Development</b>				
Parsing	High Risk	All	10/25/2025	25
Validation	Med Risk	Sami & Mahim	11/20/2025	5
Interface Development	Low Risk	Huzaifa	11/15/2025	5
Integration	Milestone	Huzaifa	11/20/2025	5
<b>Documentation</b>				
Research	On Track		10/20/2025	10
System Diagrams	On Track	Huzaifa	11/12/2025	3
<b>Testing</b>				
Integration Testing	Low Risk		11/25/2025	2
Usability Testing	Low Risk		11/28/2025	3
Issues/Bugs Fixes	Low Risk		11/25/2025	5

Figure 11: Gantt chart

Documentation activities were conducted in parallel with development throughout the project. Experimental findings, assumptions, and limitations were continuously recorded to ensure reproducibility and clarity, even when certain decoding attempts did not yield conclusive results. This parallel documentation approach was essential in a research driven project where learning outcomes are as important as final implementations.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

A testing and consolidation phase was planned toward the end of the project timeline. Given the absence of a finalized user interface and the experimental nature of the decoding tools, testing focused primarily on validation of decoding logic and consistency checks rather than traditional user facing testing.

### **9.1 Communication & Progress Monitoring**

Project planning also included regular communication with customers. Weekly meetings were scheduled as the primary mechanism for progress reporting, discussion of findings, and alignment of expectations. During these meetings, the team presented ongoing results, demonstrated experimental tools and visualizations, and discussed challenges encountered during the decoding process. It should be noted that due to scheduling constraints, not all weeks included a formal meeting.

## **10. Feedback**

This course offered a rare chance to tackle a problem that felt very close to real research and engineering work. One of its biggest strengths was the freedom to explore an open-ended task without strict, predefined specifications. That freedom encouraged independent thinking, experimentation, and critical questioning of assumptions, which are key skills in research-focused and advanced technical roles.

The project deepened our understanding of reverse engineering methods, binary data analysis, and sensor-based systems like SLAM. It showed how to make progress when documentation is missing, how to validate results without a clear ground truth, and how to structure exploratory work so that it remains systematic and reproducible. Being required to document failed attempts and limitations was especially helpful, because it highlighted learning and reasoning rather than only the final outcome.

Interaction with the customer was another major positive. Regular discussions helped clarify expectations, confirm intermediate results, and keep the work focused despite the uncertainty of the problem. This mirrored real-world research collaboration and was an important part of the learning experience.

One area for improvement would be to offer slightly clearer guidance on how research-oriented projects are evaluated, especially when a full technical implementation is not realistic. Providing a few examples of past exploratory or reverse-engineering projects could also help students better judge the expected scope and depth of their work at the start.

Overall, the course was demanding but very rewarding. It built a strong bridge between theory and practical research skills and fostered a mindset cantered on exploration, evidence-based reasoning, and technical curiosity.

Hack the dataset from X200GO - 3D modelling sensor	Muhammad Huzaifa
Final report	Date: 2025-12-22

## 11. References

- [1] D. Loss, "Binary Parsing Tools," [Online]. Available: <https://github.com/dloss/binary-parsing>.
- [2] F. AG, "Fixposition Documentation.," [Online]. Available: <https://docs.fixposition.com/>.
- [3] ". S. L. S. S. T. Stonex. [Online]. Available: <https://stonex.it/product/x200go-slam-laser-scanner>.
- [4] N. M. E. Association, "NMEA 0183 Standard.," [Online]. Available: <https://www.nmea.org>.
- [5] H. Y. Z. L. a. D. S. J. Caballero, "Polyglot: Automatic Extraction of Protocol Message Format Using Dynamic Binary Analysis," *Proc. 14th ACM Conf. Computer and Communications Security*, 2009.
- [6] Binwalk, ""Firmware Analysis Tool." [Online].," [Online]. Available: <https://github.com/ReFirmLabs/binwalk>.
- [7] "White Label SLAM Application.," [Online]. Available: <http://www.slam100.com/slam-user-guide/> .
- [8] S. A. A. M. a. M. H. S. Häkkinä, "Project Source Code Repository," [Online]. Available: <https://github.com/huzaiifaaaaa/x200-go>