Ex Mente

Python Assignment

Data Analysis Python Package

Authors:
R. Hayman

14/02/2024

Contents

1 Introduction

This project involves the design of a Python package module that can be installed and imported for use in other projects. The Python package module is focused on the processing and analytics of the data frame class type of Pandas, a statistical Python module.

Besides the data analysis module being a more encapsulating module for important statistical and plotting functions, for the Pandas data frame class, it also brings with it some extra technical advantages like multi-threading function execution with multiple data frames, as well as file syncing with the implementation of rsync.

The developed module were also used with test data to demonstrate it's functionality and advantages. Lastly, a version control system has been setup for the data analysis module on Github to ease the development and improvements of the module in different categories.

2 Implementation

2.1 Overview

2.2 Data Frame Class

2.3 Data Analysis

Figure 1 shows the setup for the practical implementation.

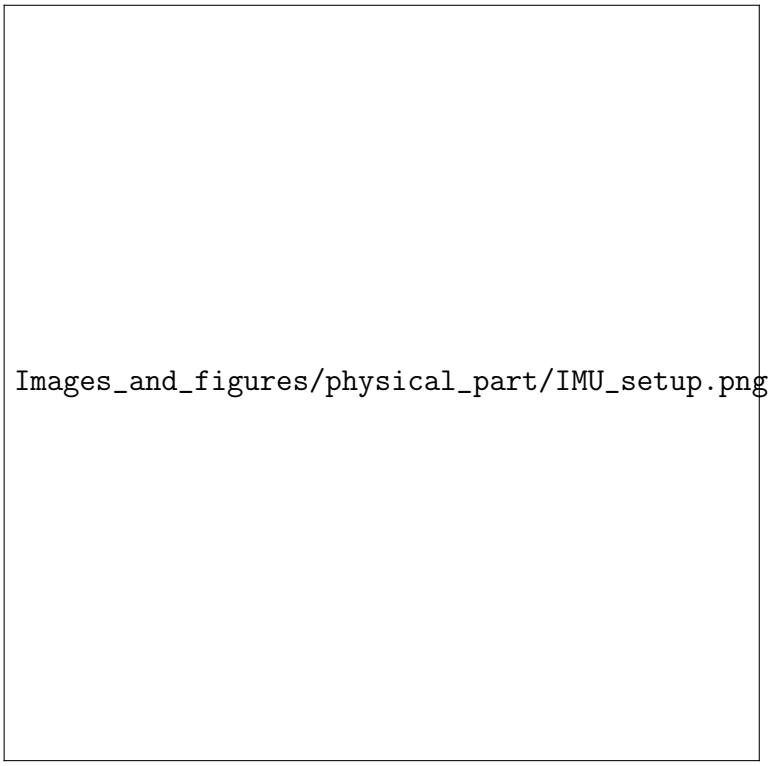

Figure 1: The setup for the practical implementation.

- A linear acceleration of ± 2g.
- A magnetic field full scale of ± 4 gauss.
- An angular rate of ± 245 dps.

3 Challenges

With the development of the data analysis Python module, some challenges were met. The 3 most prominent challenges were learning the intricacies of the Pandas module, multi-threaded function implementation with multiple data frames and lastly the implementation of rsync file syncing.

3.1 Pandas Module

The intricacies of the Pandas module

3.2 Multi-Thread Function Implementation

Multi-threaded implementation of functions will always be the more challenging parts of a project due to special care having to be taken on locking and unlocking different resources for each thread to use, in order to prevent potential problematic occurrences.

Examples of such occurrences are thread race condition and deadlocks. Thread race conditions occur when multiple threads are writing to the same memory location without there being any organisation/syncing between threads. Thus the final result will be reliant on the finishing conditions of the threads.

Another

Lastly, it should also be mentioned that debugging multi-thread programs is more challenging due to the parallel nature of the program making stepping through code more of a choir.

3.3 RSync Implementation

4 Version Control

4.1 Overview

In order to ensure that development on the Data Analysis is as progressive as possible, the Git version control system has been set in place to divide the project in clear logical development areas and to medicate potential regressions that might occur as development continues.

For this project it has been decided on, to have 3 main branches that branch of from the master branch. The 3 main branches, being development, bug-fix and documentation, each has their respective purpose as explained in sub-subsection 4.1.1, 4.1.2 and 4.1.3, but all 3 main branches are permanent branches that will still be kept after their changes has been merged with the master branch.

Sub-branches can be created from any of the 3 main branches but these branches will only be temporally, meaning that they will be deleted after their changes has been merged with the respective main branch. A visual representation of the Git repository branch structure can be seen in Figure 2

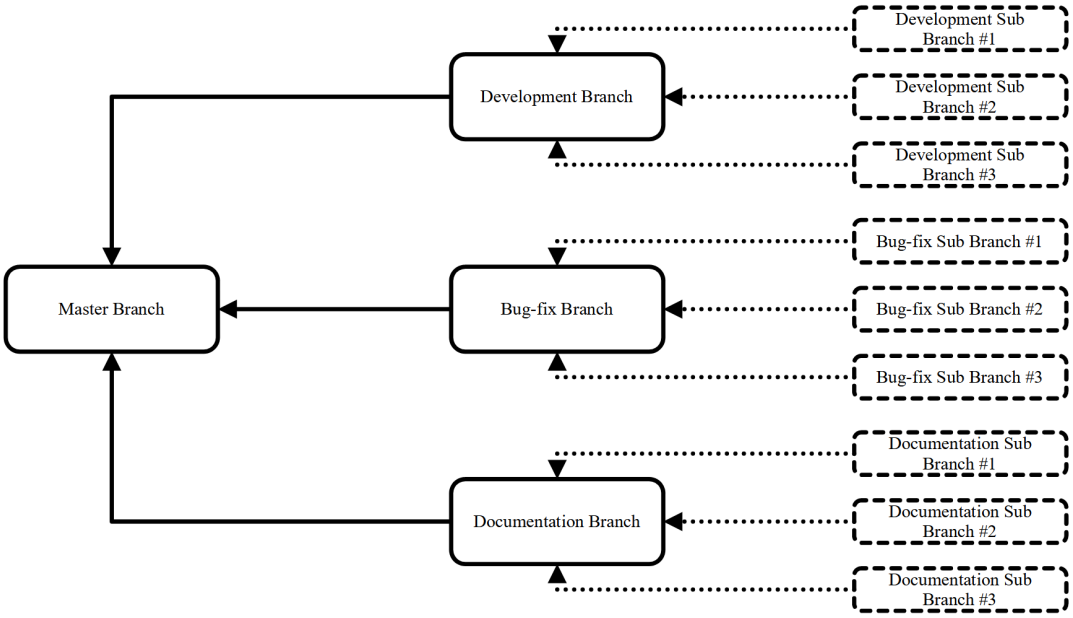

Figure 2: The Git repository branch structure for the project

4.1.1 Development Branch

The main purpose of the development branch is to sandbox new features/functions being developed from the master branch. This strategy prevent the new features from potentially causing bugs/regressions in the master branch.

4.1.2 Bug-fix Branch

The main purpose of the bug-fix branch is to fix bugs that have been discovered, on existing features of the Data Analysis package. These fixes are then tested in the bug-fix branch before being merged with the master branch.

4.1.3 Documentation Branch

The purpose of the documentation branch is to update the package manual and doc strings comments of each component of the package. This means that all changes made in this branch should not in any way alter the functionality of the package, when compared compared to the master branch.

4.2 Examples

5 Data Analysis

5.1 Dataset

5.2 Analysis on Dataset

5.3 Multi-Thread Advantage

6 Conclusion

Overall the project was a success, with the core functionality of the data analysis package being fully developed with the inclusion of multi-threading functionality, plotting capability and a file sync system. The functionality of the package has also been quite thoroughly tested with test data.

Lastly, a git repository has been setup to organise the development areas of the data analysis package in a logical manner and to prevent potential regressions that might occur with future development.

Some improvements that could be made in the future to the data analysis package a more robust plotting module that has more features/plotting options. The plotting time could then also be further enhanced by integrating multi-threading into the plotting functions.

References

[1] S. Ferrari and A. Vaghi, "Demining sensor modeling and feature-level fusion by Bayesian networks," in IEEE Sensors Journal, vol. 6, no. 2, pp. 471-483, April 2006, doi: 10.1109/JSEN.2006.870162.

[2] J. K. Wu and Y. F. Wong, "Bayesian Approach for Data Fusion in Sensor Networks," 2006 9th International Conference on Information Fusion, 2006, pp. 1-5, doi: 10.1109/ICIF.2006.301810.

[3] Manon Kok, Jeroen D. Hol and Thomas B. Schon on (2017), "Using Inertial Sensors for Position and Orientation Estimation", Foundations and Trends in Signal Processing: Vol. 11: No. 1-2, pp 1-153. http://dx.doi.org/10.1561/2000000094

[4] A. Assa and F. Janabi-Sharifi, "A Kalman Filter-Based Framework for Enhanced Sensor Fusion," in IEEE Sensors Journal, vol. 15, no. 6, pp. 3281-3292, June 2015, doi: 10.1109/JSEN.2014.2388153.

[5] L. Schäffer, Z. Kincses and S. Pletl, "A Real-Time Pose Estimation Algorithm Based on FPGA and Sensor Fusion," 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), 2018, pp. 000149-000154, doi: 10.1109/SISY.2018.8524610.

[6] M. Ghanbari and M. J. Yazdanpanah, "Delay Compensation of Tilt Sensors Based on MEMS Accelerometer Using Data Fusion Technique," in IEEE Sensors Journal, vol. 15, no. 3, pp. 1959-1966, March 2015, doi: 10.1109/JSEN.2014.2366874.

[7] STMicroelectronics, 2021. STEVAL-MKI159V1 - STMicroelectronics. [online] STMicroelectronics. Available at: ¡https://www.st.com/en/evaluation-tools/steval-mki159v1.html¿ [Accessed 27 September 2021].

[8] STMicroelectronics, 2021. STM32F429/439 - STMicroelectronics. [online] STMicroelectronics. Available at: ¡https://www.st.com/en/microcontrollers-microprocessors/stm32f429-439.html¿ [Accessed 27 September 2021].