**Registration Benchmark data\_processing.py Manual**

1. Program Overview

The ‘data\_processing.py’ script is used to preprocess point cloud data. This script contains functions to preprocess point cloud data from their source directory and calculates overlap percentages. It handles both 'sun3d' and 'eth' dataset types. It also includes utilities to draw registration results and calculates total points in a pair of point clouds.

2. Function Descriptions

Main Function:

preprocess\_point\_clouds(dataset\_name: str, overlap\_threshold: float = 0.4) -> None

This main function preprocesses the point cloud data in a directory and calculates overlap percentages. Determines the dataset type and calls the appropriate preprocessing function.

Auxilliary Functions:

preprocess\_sun3d\_point\_clouds(dataset\_name: str, overlap\_threshold: float = 0.4) -> None:

Preprocesses the SUN3D point cloud datasets and calculates overlap percentages between pairs of point clouds. It reads the point cloud pairs from a log file, calculates the overlap percentage, and saves it in a matrix. Pairs of point clouds with an overlap percentage above a certain threshold are saved to a text file, along with their corresponding ground truth transformation matrix.

preprocess\_eth\_point\_clouds(dataset\_name: str, overlap\_threshold: float = 0.4) -> None:

Preprocesses the ETH point cloud datasets and calculates overlap percentages between pairs of point clouds. The overlap percentage is computed for each pair of point clouds and saved in a matrix. Pairs of point clouds with an overlap percentage above a certain threshold are saved to a text file.

Note: The overlap percentage is a measure of how similar two point clouds are. The higher the overlap percentage, the more similar the point clouds are. This is important in many applications, such as 3D object recognition, where it is necessary to compare and match different point clouds.

3. Imports and Dependencies

* `open3d` as `o3d`: A modern library for 3D data processing.
* `os`: Provides functions for interacting with the operating system.
* `numpy` as `np`: The fundamental package for scientific computing with Python.
* `pathlib.Path`: Represents the system’s path.
* `utils.draw\_registration\_result`: A utility function for drawing the registration result.
* `utils.overlap\_percentage`: A utility function for calculating the overlap percentage.
* `re`: Provides regular expression matching operations.
* `pandas` as `pd`: A data analysis and manipulation tool.
* `utils.calculate\_total\_points`: A utility function for calculating the total points in a pair of point clouds.
* `copy`: A module for generic shallow and deep copy operations.

4. Operation/Usage

Invoke the main function preprocess\_point\_clouds with the dataset's name and an optional overlap threshold. Based on the dataset name, it will call either the preprocess\_sun3d\_point\_clouds or preprocess\_eth\_point\_clouds function.

5. Output and Interpretation

The functions preprocess\_sun3d\_point\_clouds and preprocess\_eth\_point\_clouds generate an overlap matrix CSV file and a text file containing overlapping point cloud pairs if they exceed the overlap threshold.. A "Preprocessing complete" message will indicate a successful operation.