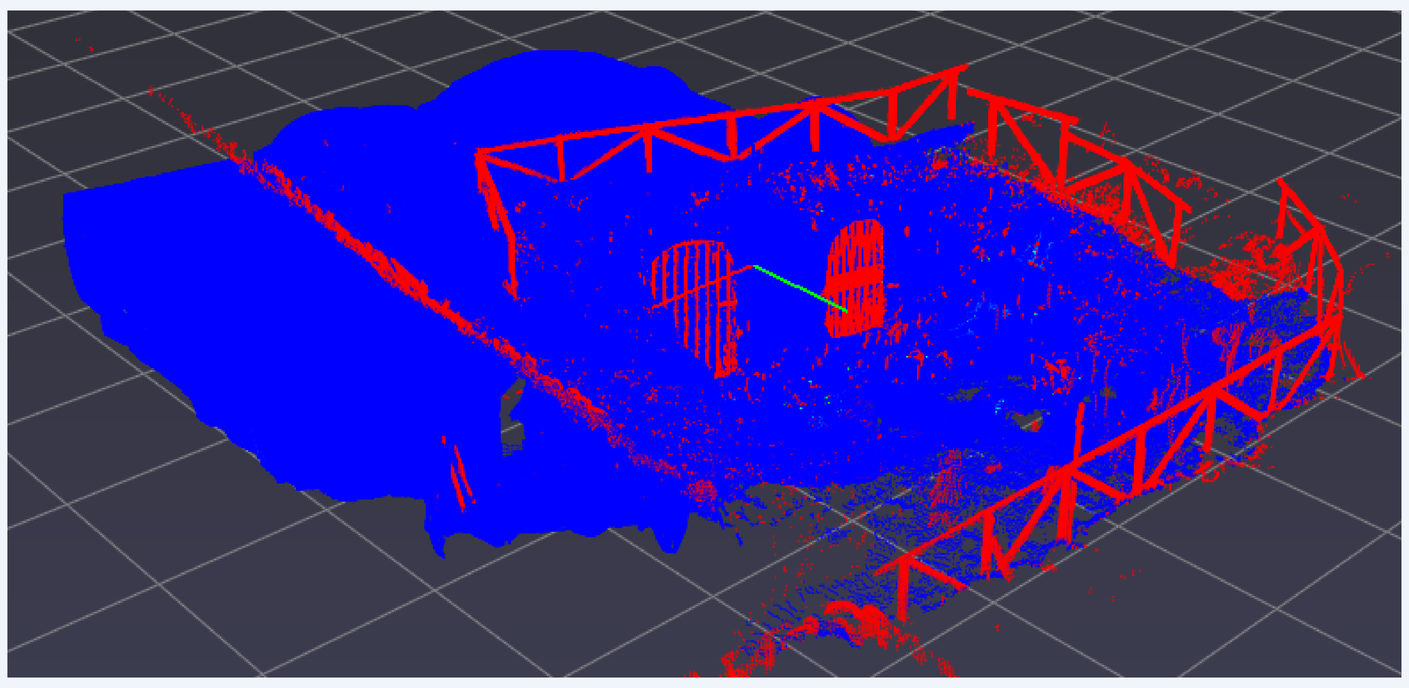
Semantically Meaningful Clustering of Cultural Heritage Sites

Background

Laser scans of Cultural Heritage (CH) sites produce point clouds – sets of 3D data points – which can be visualised as 3D models. However, these scans contain unwanted information that one must remove through point cloud cleaning - a mainly manual, time-consuming process. It must be automated and made more efficient.

Objectives

1. Investigate whether clustering methods produce more semantically meaningful clusters than k-means, through point feature clustering on point clouds with varying dimensionality.
2. Determine the upper bound accuracy when using clusters to perform binary classification on point clouds.
3. Simplify and accelerate point cloud cleaning.



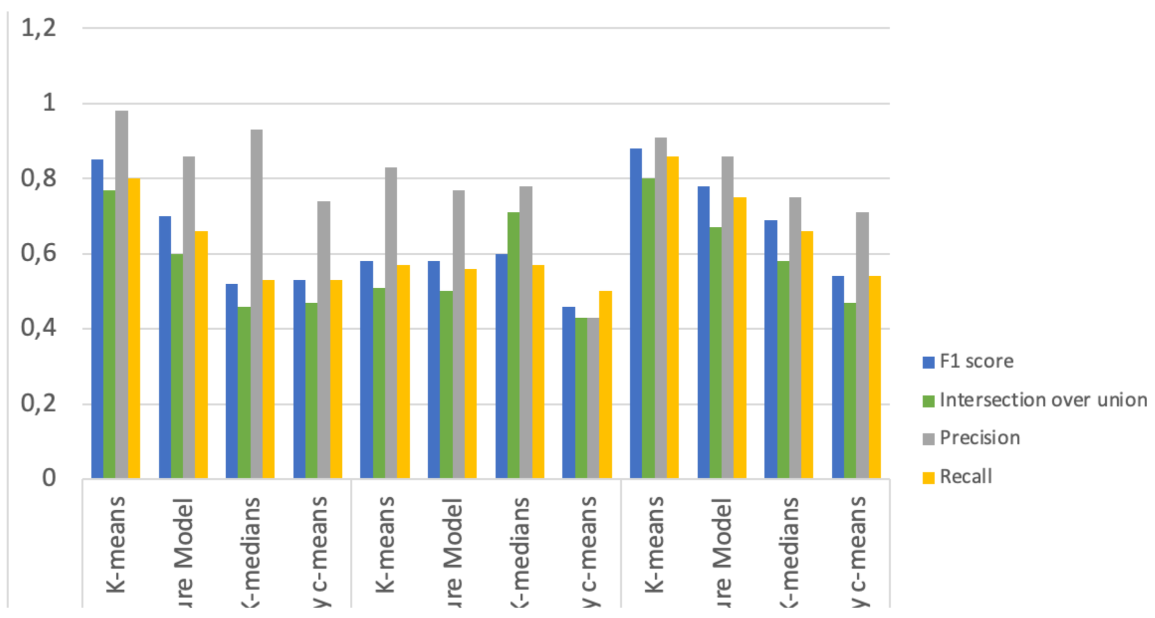
Ground truth labels - Church Point Cloud

Clustering

We explore the objectives through different classes of clustering algorithms and compare against k-means

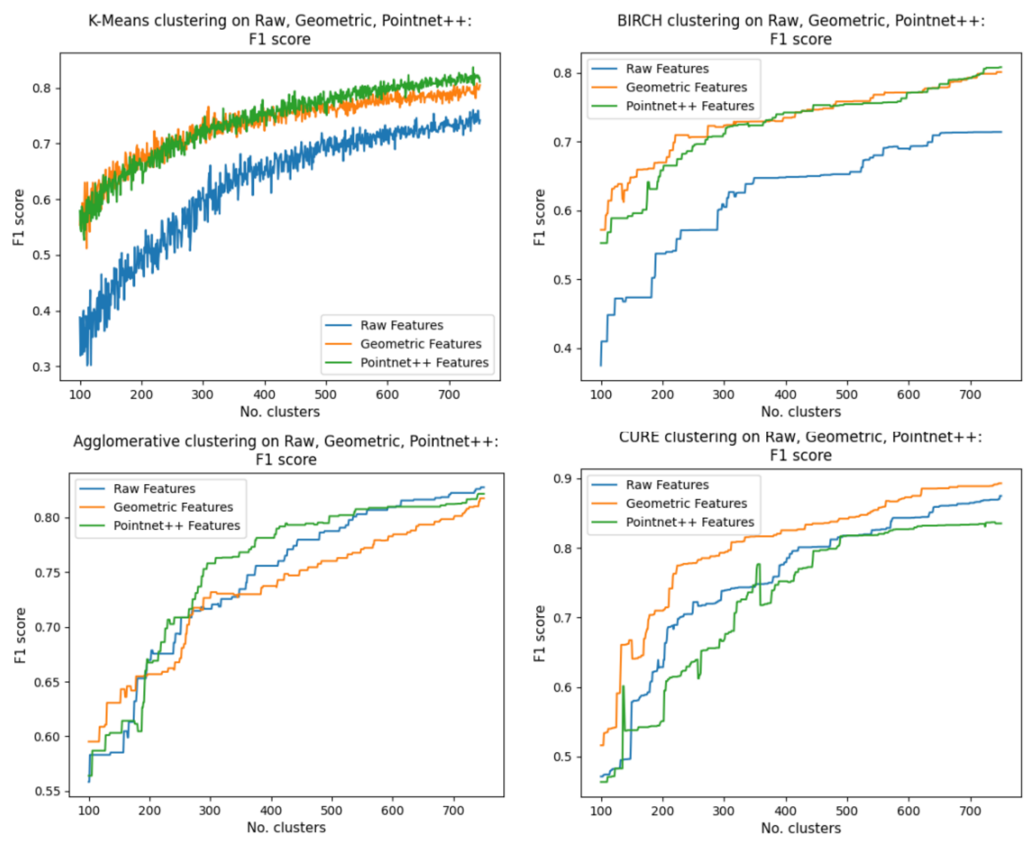
Partitional + GMM

- Implement: GMM, K-medians and Fuzzy c-means.
- GMM maintains a strong performance in the clustering and classification task. throughout all three datasets.
- GMM produces a small amount of clusters that can simplify and accelerate the cleaning task.
- The performance of all the algorithms is optimal when utilising the PointNet++ dataset.
- Encoding additional features into the dataset is valuable.



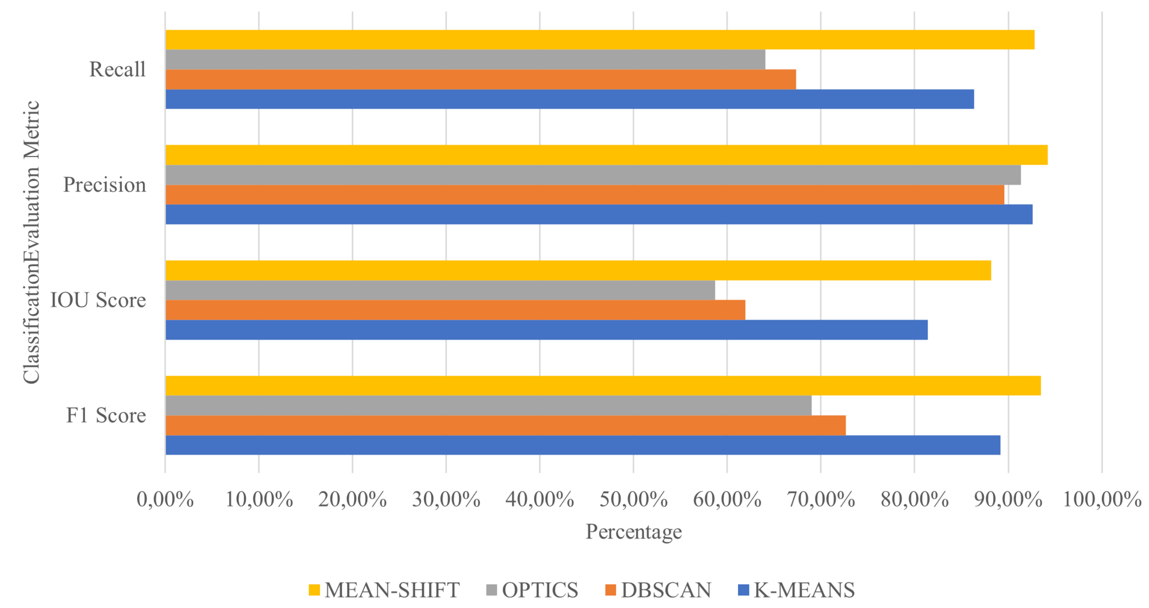
Hierarchical

- Implement: Agglomerative, BIRCH, CURE clustering.
- All algorithms match or exceed K-means in many clustering and classification metrics.
- CURE and Agglomerative show the most promising results.
- High time and space complexity.
- Algorithms perform well with low dimensionalities and down sampled data. This reduces the effects of increased complexity.



Density

- Implement: DBSCAN, OPTICS, Mean-Shift.
- Mean-Shift accurately classifies point cloud into *keep* and *discard* and outperforms k-means on two datasets.
- Algorithms produce too many clusters, despite good classification results.
- DBSCAN and OPTICS perform sub-optimally.
- Feature selection affects results.
- Best classification results achieved on raw dataset.



Conclusions

Clustering can classify CH point clouds into *keep* and *discard* clusters. Some methods are impractical for point cloud cleaning because they produce numerous clusters. Feature set dimensionality and selection affects clustering algorithms' performance. Cluster validity metrics are insufficient for representing the clusters' usefulness for this task.

University of Cape Town
Department of Computer Science
dept@cs.uct.ac.za
021 650 2663

Project Team
Leah Gluckman glclea001@myuct.ac.za
Jemma Sundelson sndjem001@myuct.ac.za
Jared May myxjar002@myuct.ac.za

Supervisors
A/Prof. Patrick Marais
Luc Hayward