Semantically Meaningful Clustering of Cultural Heritage Sites

Background

Constructing 3D models from laser scans of Cultural Heritage (CH) sites is vital for CH preservation. These scans produce point clouds - a set of data points in 3D space - which can be visualised as 3D models. However, the construction process is challenging because laser scans of CH sites contain unwanted information, such as foliage, people and scaffolding, that do not belong to the CH site and which one must remove. Removing this noise is called point cloud cleaning - a predominantly manual, laborious and time-consuming task which must be automated and made more efficient.

Objectives

- Investigate the decomposition of point clouds into semantically meaningful clusters through point feature clustering on datasets with varying dimensionality.
- Simplify and accelerate point cloud cleaning and assess whether clustering can augment the classification process.

Research Questions

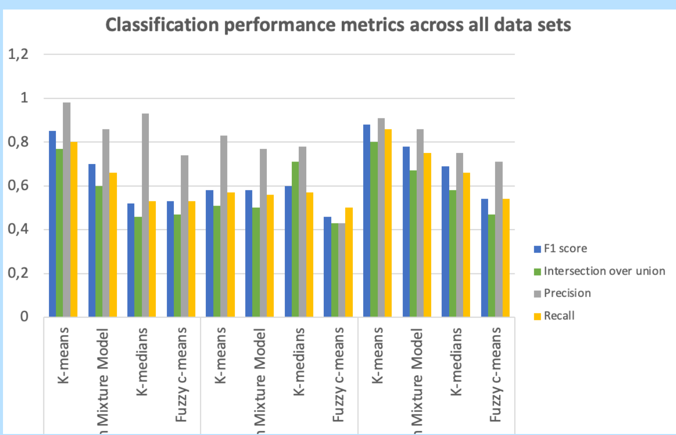
- Do clustering methods produce more semantically meaningful clusters on CH sites point clouds than k-means, using cluster validity metrics?
- What is the upper bound accuracy when using the resulting clusters to perform binary classification on a point cloud?

Clustering

Each sub-project explores the objectives through several different classes of clustering algorithms and uses k-means clustering as a baseline clustering algorithm for comparison.

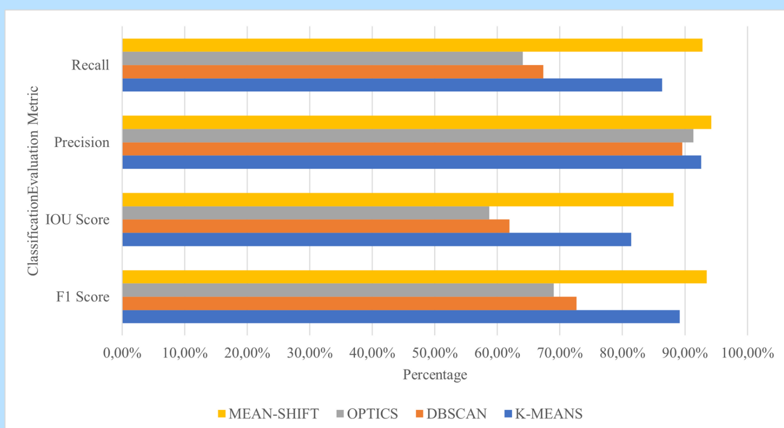
Partitional + GMM

Partitional clustering assigns a set of data points into k-clusters by using iterative processes. A Gaussian Mixture Model (GMM) is a soft clustering algorithm that assigns various probabilities to points. We implement GMM, K-medians and fuzzy c-means. GMM maintains strong performance in the clustering and classification task throughout all three datasets. Additionally, the performance of all the algorithms enhances when utilising the PointNet++ dataset. Therefore, encoding additional features into the dataset is valuable.



Density

Density-based clustering methods can identify arbitrarily shaped clusters and effectively discover noise. We implement DBSCAN, OPTICS and mean-shift. Mean-shift can accurately and efficiently partition a CH point cloud into clusters that can classify as keep and discard. It outperforms k-means in classification evaluation metrics on two datasets but produces too many clusters. Thus, it is impractical for point cloud cleaning. DBSCAN and OPTICS produce poor classification results and do not produce more semantically meaningful clusters than k-means.

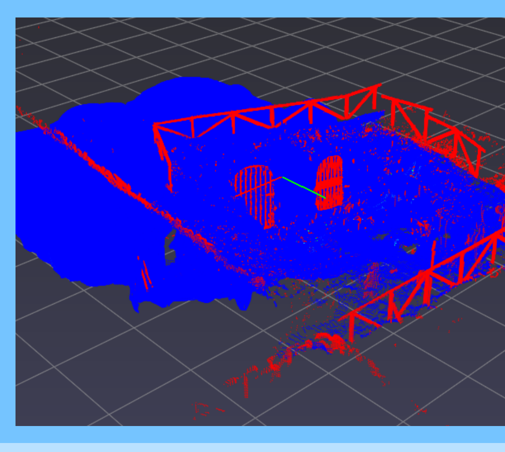


Hierarchical

Hierarchical clustering algorithms build hierarchical clusters by merging (agglomerative methods) or dividing (divisive methods) clusters successively. This hierarchy of clusters is represented as a tree (or dendrogram). The root of the tree structure is the unique cluster that gathers all the samples, the leaves being the clusters with only one sample.The 3 hierarchical clustering algorithms explored were Agglomerative clustering, BIRCH Clustering and CURE Clustering.

Conclusions

Clustering methods can classify CH point clouds into *keep* and *discard* clusters. However, they may not be a viable solution to point cloud cleaning because they produce numerous partitions rather than one for each classification label. Feature set dimensionality and selection affects clustering algorithms' performance. In addition, the cluster validity metrics are insufficient for representing the usefulness of the produced clusters. Thus, selecting the optimal algorithm is complex. Despite this, GMM and mean-shift outperform the baseline k-means algorithm and produce clusters that, when joined for classification, resemble the actual ground truth labels.



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/Professor
Patrick Marais
Luc Hayward