The Journal of Open Source Software

pymccrgb: Color- and curvature-based classification of multispectral point clouds in Python

Robert Sare[1] **and George E. Hilley**[1]

1 Department of Geological Sciences, Stanford University

DOI: 10.21105/joss.01777

Software

- Review
- Repository
- Archive

Submitted: 24 September 2019
Published: 31 October 2019

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC-BY).

Summary

Digital elevation data are used extensively in the geophysical sciences, including conventional light detection and ranging (lidar) point clouds and very high density photogrammetric datasets produced from drone surveys or stereo imagery. Classifying ground and vegetation points is an important step in topographic data analysis in geomorphology and environmental science, especially as many datasets increasingly image vegetation and other subtle features in fine detail. `pymccrgb` is a package for classification of point cloud data with point colors or other multispectral information, providing a simple interface for point classification to scientific Python users.

This package offers Earth scientists studying surface processes or hazards an efficient method for extracting ground points from unclassified point clouds, and may also be useful for detailed vegetation height measurements in forestry or rangeland ecology. The method's implementation uses Numpy, scikit-learn, and PDAL and it is intended to be easy to extend to other supervised classification methods or point classes (PDAL Contributors, 2018; Pedregosa et al., 2011; van der Walt, Colbert, & Varoquaux, 2011). The core functionality builds on a popular open source algorithm, multiscale curvature classification (MCC) (Evans & Hudak, 2007), by training a support vector machine classifier using color features of vegetation points and updating ground classified points according to color similarity.

This new two-stage algorithm, MCC-RGB, requires fewer iterations than the MCC method and removes low vegetation points in settings that can challenge MCC. Users can choose to re-classify ground points in a single step or at user defined height ranges representing multiple vegetation classes. The package provides a Python API to both methods.

Acknowledgements

Example data in this package was collected by the National Center for Airborne Laser Mapping and hosted by the OpenTopography facility with support from the National Science Foundation under NSF award numbers 1557484, 1557319, and 1557330. We would like to thank Daniel Livingston and Martí Bosch for their constructive reviews and editor Kristen Thyng for her editorial handling of the manuscript.

References

Evans, J. S., & Hudak, A. T. (2007). A multiscale curvature algorithm for classifying discrete return LiDAR in forested environments. *IEEE Trans. Geosci. Remote Sens.*, *45*(4), 1029–1038. doi:10.1109/TGRS.2006.890412

PDAL Contributors. (2018). PDAL Point Data Abstraction Library. doi:10.5281/zenodo.2556738

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A structure for efficient numerical computation. *Computing in Science and Engineering*, *13*(2), 22–30. doi:10.1109/MCSE.2011.37