Classification of High-Dimensional Rasterized LiDAR Datasets Using Evolutionary Computation

Henrique Momm¹, Evan Snapp²

¹Middle Tennessee State University, ²Computer Science, Middle Tennessee State University

LiDAR datasets are commonly represented as irregularly spaced three-dimensional point clouds. To facilitate feature classification and data manipulation, point clouds are often rasterized into regular grids as multi-channel images. Modern LiDAR sensors record a large number of measurements for individual laser points. Therefore, image channels can be generated from single measurements, combination of two or more measurements, spectral transformations, and convolution operators. Depending on the number of measurements and selection of operators (and their controlling parameters) often LiDAR datasets are converted into images with hundreds of channels. The selection of which channels are needed for classifying specific features is an enormous task. Alternatively, heuristics computer algorithms can be used to expedite this process. In this study, genetic programming was evaluated in reducing data dimensionality specifically for the task of LiDAR images classification. Using an image composed of 95 channels, the accuracy and convergence of the algorithm were evaluated in classifying multiple features. Genetic programming was designed to generate candidate solutions represented by computer programs in the form of mathematical equations (spectral indices) in a learn-from-examples fashion. Given the size of the search space compounded with the randomness involved in the creation of the initial guess from which the algorithm begins, it is possible that the algorithm is unable to converge to the global minimum. Three techniques were investigated to address these limitations. Our findings indicate that introducing new candidate solutions at key points during the evolutionary process increased the diversity of the candidate solution set and by extension the chances of improved overall performance of the algorithm. The final solutions are mathematical expressions representing spectral indices composed of a small subset of the image channels signifying the selection of key channels needed to classifying the feature of interest. The findings of this study support image processing in a wide range of applications faced with the challenge of converting high-dimensional data into actionable information.