Automated building detection and mapping from simultaneously acquired airborne hyperspectral and lidar data with Mask R-CNN

Øivind Due Trier, Norwegian Computing Center, Section for Earth Observation, Gaustadalléen 23A, P.O. Box 114 Blindern, NO-0314 Oslo, Norway.

Urban maps in Norway are currently updated using manual photo interpretation on stereo aerial imagery. However, there is often a substantial delay after completion of construction work until new buildings, roads, etc. appear in updated versions of the urban maps.

Automated pixel-based urban land cover classification from multispectral aerial images of very high resolution is difficult since the same spectral values may occur within several land cover types. Airborne hyperspectral data may provide better discriminative power. However, there is still the problem that the same types of material may exist within different land cover types, such as buildings, roads, parks, gardens, etc.

With the rapid development in deep neural network methods and computer processing resources, it should be possible to develop methods that could automate at least some of the urban map revision tasks. Detection and mapping of new and/or changed buildings is one such task that is important in Norway.

Airborne hyperspectral and lidar data were acquired simultaneously for an urban/suburban area in Bærum municipality near Oslo, Norway. The hyperspectral data had 30 cm pixel spacing and 186 channels in the visible and near-infrared spectrum. The lidar data had five emitted pulses per m^2. The data was georeferenced to UTM zone 32 N. Each lidar (x, y, z) point was classified as either 'ground' or 'other'. Atmospheric correction of the hyperspectral data was not considered necessary, as the data was collected within a few hours on a single day.

The deep neural network Mask R-CNN was used for automated building detection.

The combined data were split into three separate geographic regions named training, validation and test. The training data was used to tune the internal parameters in the deep neural network. The validation data was used to select the best set of parameters encountered so far. The test data was used for a final quantification of the performance of the deep neural network on data not seen during training.

Mask R-CNN was able to detect new buildings and changed buildings, compared to the vector data of 2018, which was the main purpose. For quantitative evaluation, detection of existing buildings was used. False negatives, i.e., true buildings that were not detected by Mask R-CNN, were observed for small buildings completely covered by tree canopies. False positives, i.e., locations without buildings that were falsely detected as buildings by Mask R-CNN, were infrequent. For building detection, Mask R-CNN performed better than a simple thresholding on object height and NDVI. However, Mask R-CNN was not able to map the exact boundaries of the buildings.

For the purpose of automated map revision, Mask R-CNN was shown to be useful to flag new and changed buildings. However manual confirmation and exact delineation by a human operator was still needed.

Acknowledgments: This research is part of a project on machine learning in map revision, financed by: Regional Research Fund Viken, Bærum municipality, TerraTec AS and Geovekst.