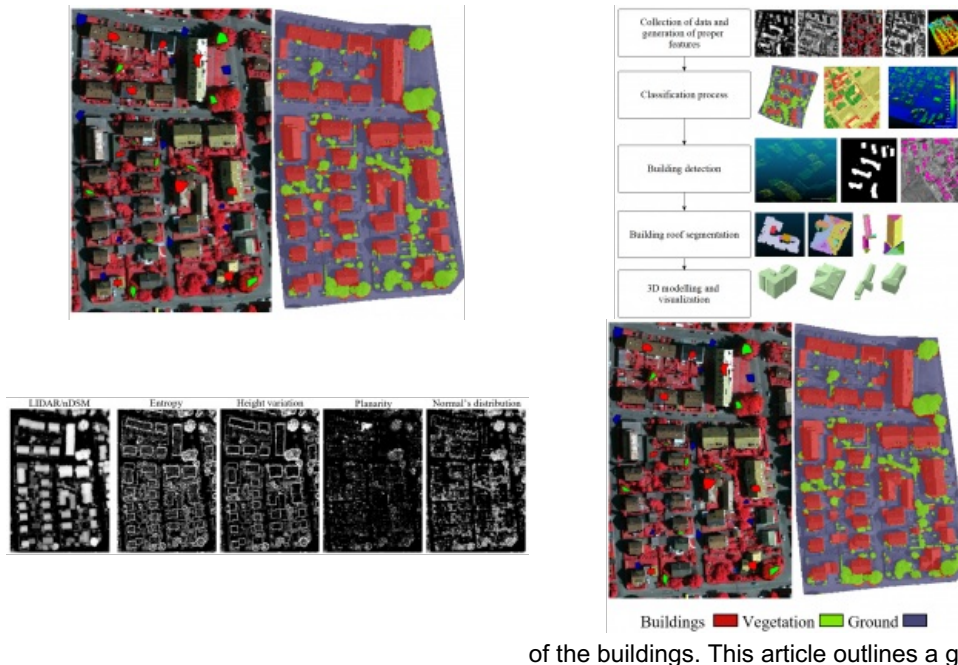


# A GENERIC WORKFLOW FOR AUTOMATIC BUILDING DETECTION AND 3D MODELLING

## Advanced urban 3D modelling and visualization



This article outlines a generic workflow for automatic building detection and 3D modelling using modern technologies, to support applications ranging from urban planning and cadastre to change detection and navigation.

The automatic detection, data extraction, 3D modelling and visualization of buildings in urban areas using remote sensing data is an essential task in various applications such as cadastre, urban and rural planning, change detection, mapping, updating geographic information systems, monitoring, housing value and navigation. Even today, this task remains challenging due to the inherent artefacts (e.g. shadows) in the remote sensing data used, as well as the differences in viewpoints, surrounding environment and complex shape and size

of the buildings. This article outlines a generic workflow using modern technologies.

Besides the recent developments in image processing, advances in computer vision have promoted automated methods able to generate precise 3D models from overlapped multiple 2D imagery data derived from aerial platforms. Such methods apply a dense image matching (DIM) algorithm which extracts a textured dense 3D point cloud of a region or an object of interest. DIM is an affordable process compared to other approaches that use other types of sensorial data such as Lidar. In this area, numerous robust stereo image matching algorithms have been developed, each of which has its own advantages and limitations. A generic workflow for building detection and 3D modelling includes the following steps: i) collection of data and generation of proper features, ii)

classification process, iii) building detection, iv) building roof segmentation, and v) 3D modelling and visualization.



Generic workflow for building detection and 3D modelling.

### Generation of additional features

Depending on the data source employed, building detection techniques can be classified into three groups: (i) ones that use airborne or satellite imagery data, (ii) ones that exploit three-dimensional information, and (iii) those that combine both data sources. However, the two main limitations of using information from multi-modal sources (e.g., Lidar and imagery data) are the additional cost of acquisition and processing, and the issues related to co-registration. For this reason, in real-life applications such as the cadastral ones, sometimes only one type of data is considered. To this end, several indices and features are calculated to efficiently distinguish buildings from the other urban objects such vegetation and ground. In this context, depending on the data used, the normalized difference vegetation index (NDVI) is calculated (when the NIR band is available in images) and the normalized digital surface model (nDSM) is calculated (when DIM or Lidar point clouds are available). However, additional features can be calculated and image-stacked, especially from Lidar point clouds, to

further contribute to the classification performance. Such features come from a physical interpretation of the information, e.g. the entropy, the height variation, the planarity and the distribution of the normal vectors.



Generation of additional features from Lidar point clouds.

## Classification and building detection

Usually, the methods of building detection are discriminated into the ones that apply a supervised machine-learning scheme and those that use a model-based approach. The main advantage of the machine learning approaches is that they are flexible and data-driven methods, requiring only training samples to successfully generalize the building properties and thus to perform an accurate classification. In contrast, model-based approaches consist of many parameters that need to be fine-tuned for each study area. Therefore, supervised learning paradigms provide higher generalization capabilities, i.e. robustness against data being outside the training set. Recently, in the context of machine learning, state-of-the-art algorithms like deep learning classifiers through convolutional neural networks (CNNs) have been efficiently applied for the building detection task.

In general, a CNN classifier has two main components: the convolutional layer and the classification layer. A convolutional layer is essentially a network feature extractor that employs convolution filters (i.e. transformations) to the input data (image-stack features). These extracted network features are able to optimize the classification performance. Spatial coherency is an important element of the transformations involved in the convolutional layer. This is an important property of a deep CNN model since spatial characteristics significantly affect building detection accuracy. The aim of the classification layer is actually a supervised learning scheme with the capability of transforming the inputs from the convolutional layer into desired outputs, i.e. the labelled classes. Therefore, a CNN classifier, in contrast to a shallow machine learning method, first filters the input data in a way to maximize the classification accuracy and then performs the classification. The output of the CNN is a classified image on a pixel level, including information associated with the label of each class. Post-morphological processing is adopted to reduce classification noise, taking into consideration the spatial coherency property, i.e. through min operators followed by majority voting filters, etc. Finally, to evaluate the final building detection results, objective criteria are used such as the completeness, correctness and quality rates based on the TP, FP and FN entities, whereby TP stands for true positives (e.g. reference building pixels that were correctly detected), FP stands for false positives (e.g. building pixels that not exist in the reference dataset) and FN stands for false negatives (e.g. reference building pixels that were not detected).



Collection of training samples for each class (left) and classification results through a CNN classifier (right).

## Building roof segmentation and 3D modelling results

The extracted building boundaries from the classification process are slightly dilated in order to clip the raw Lidar or DIM point cloud. Then, for each 3D point cloud of each building, a building roof segmentation process is carried out. The most-used plane detection techniques from 3D point clouds are region growing, RANSAC and Hough methods. In fact, adaptive point randomized Hough transform (RHT) can extract satisfactory results, satisfying greatly the accuracy vs. computational time trade-off. For each detected plane, the corresponding boundaries are extracted to generate the associated 3D polygons. Once the normalized height values of each polygon vertex are available, the corresponding 3D building model can be extracted.



Building roof segmentation and 3D modelling.

## Conclusion

Automatic building detection and 3D modelling is a continuous, essential and crucial task for a variety of applications. Modern technologies support the development of a generic workflow. Two key emerging technologies are: i) various new sensors that can provide multiple information (e.g. multi/hyperspectral Lidar point clouds), and ii) cutting-edge methods such deep machine learning schemes.

### Further reading

- Maltezos, E., Doulamis, A., Doulamis, N., Ioannidis, C., 2019. Building extraction from LIDAR Data applying deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters (GRSL)*, Vol. 16, pp. 155-159.
- Maltezos, E., Doulamis, N., Doulamis, A., Ioannidis, C., 2017. Deep convolutional neural networks for building extraction from orthoimages and dense image matching point clouds. *Journal of Applied Remote Sensing*, Vol. 11, 4, pp. 042620-1-042620-22.