

1 Data

The available data are 1024x1024 images of the two United States cities Jacksonville in Florida and Omaha in Nebraska taken from the US3D Dataset that was partially published to provide research data for the problem of 3D reconstruction (Bosch et al. 2019). The images for each recorded area cover one square kilometer and can be divided into four categories with the first one being optical satellite images with three channels (RGB). Secondly visible and near infrared satellite images with eight channels (VNIR). Thirdly digital surface models (DSM). And lastly semantic labeling with five different categories.

The optical images were taken by the WorldView-3 satellite of Digital Globe from 2014 to 2016 and contain seasonal and daily differences in vegetation and sun positions. Each pixel of an image is described by three bytes representing the intensity of either red, green or blue.

Also collected by WorldView-3 were the VNIR images which contain eight channels for eight different bands of the spectrum with a ground sample distance of 1.3 meters. These images were taken over the course of all twelve months making them usable for training models that can handle seasonal appearance differences which are even more distinct than in the RGB data because certain wavelengths capture shadows and vegetation especially well. Overall this data offers more detail than the three channel RGB pictures. The eight channels of the imagery correspond to the following wavelengths:

- Coastal: 400 - 450 nm
- Blue: 450 - 510 nm
- Green: 510 - 580 nm
- Yellow: 585 - 625 nm
- Red: 630 - 690 nm
- Red Edge: 705 - 745 nm
- Near-IR1: 770 - 895 nm
- Near-IR2: 860 - 1040 nm

The given DSMs were collected using light detection and ranging technology (Lidar). They have a single channel that describes the height of each pixel with a greater number representing a higher distance to the ground.

Lastly there are semantic labeled pictures with one channel of a single byte encodes one of five different topographic classes. Those classes are vegetation, water, ground, building and clutter. The semantic labeling was done automatically from lidar data but manually checked and corrected afterwards.

For all four categories of data the area covered in a single image is one square kilometer and they contain a lot of oblique view of buildings, often with sunshine casting good shadows making the data ideal for training models that should detect them.

2 Experiments

References

Bosch, Marc et al. (2019). “Semantic Stereo for Incidental Satellite Images”. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, pp. 1524–1532.