1. Detailed steps involved in the generation of Fig. 3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Image | nDSM | | MABI | |
| Feature | Result | Feature | Result |
| High-rise buildings |  |  |  |  |  |
| Mid-rise buildings |  |  |  |  |  |
| Low-rise buildings |  |  |  |  |  |
|  | (a) | (b) | (c) | (d) | (e) |

**Fig. 3.** Typical examples of building detection using optical stereo images (Ziyuan-3): (a) image scenes for high-, mid-, and low-rise buildings; (b) and (d) represent feature images for an nDSM and the multi-angular built-up index (MABI), respectively; and (c) and (e) represent the building detection results obtained using the nDSM and MABI, respectively., i.e., pixels with nDSM (MABI) larger than a manual threshold were extracted as buildings.

1. In column (a), three true color examples in format of RGB are manually selected from a Ziyuan-3 optical stereo image.
2. In column (b), the normalized Digital Surface Model (nDSM) was generated by a top-hat morphological operation on Digital Surface Model (DSM), and then, pixels with nDSM larger than a manual threshold was extracted as buildings (i.e., see column (c)). The detailed step to generate DSM, nDSM, and the build footprint can be referred to our previous study Liu et al. (2017).
3. Multi-Angular Building Index (MABI) was calculated using nadir (NAD), forward (FWD), and backward (BWD) panchromatic images of ZY-3 satellites:

where , and are the reflectance of nadir NAD, FWD and BWD images, respectively. Pixels with larger than a manual threshold was extracted as buildings. More details can be seen in Liu et al. (2019).

Reference:

Liu, C., Huang, X., Wen, D., Chen, H., Gong, J., 2017. Assessing the quality of building height extraction from ZiYuan-3 multi-view imagery. Remote Sensing Letters, 8(9):907-916

Liu, C., Huang, X., Zhu, Z., Chen, H., Tang, X., Gong, J., 2019. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: Analysis of 45 global cities. Remote Sensing of Environment, 226: 51-73.

1. Detailed steps involved in the generation of Fig. 4

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Image | Harris | | PanTex | | MBI | |
| High-rise buildings |  |  |  |  |  |  |  |
| Mid-rise buildings |  |  |  |  |  |  |  |
| Low-rise buildings |  |  |  |  |  |  |  |
| Small buildings |  |  |  |  |  |  |  |
| Dense buildings |  |  |  |  |  |  |  |
|  | (a) | (b) | (c) | (d) | (e) | (f) | (g) |

**Fig. 4.** Typical examples of building detection using a single optical image: (a) image scenes for high-rise, mid-rise, low-rise, and small and dense buildings; (b), (d), and (f) represent the feature images for the Harris detector, PanTex index, and the MBI, respectively; (c), (e), and (g) represent the building detection results of the Harris detector, PanTex index, and the MBI, respectively, i.e., pixels with feature value larger than a manual threshold were extracted as buildings.

1. In column (a), five true color examples in format of RGB are manually selected from a Ziyuan-3 optical image.
2. For column (b) and (c), the Harris corner response of each pixel was calculated using Harris detector:

where and are the derivatives of the image in horizontal and vertical direction, is the Gaussian function, is the convolution operation, and and are the determination and trace of the matrix. Pixels with Harris corner response larger than a threshold was extracted as buildings.

1. For column (d) and (e), The Gray-Level Co-occurrence Matrix (GLCM) was calculated for each pixel in 10 combinations of directions and displacements where the shift in horizontal and vertical direction is (1, -2), (1, -1), (2, -1), (1, 0), (2, 0), (0, 1), (1, 1), (2, 1), (0, 2), and (1, 2), respectively. For each GLCM, the contrast metric was calculated:

where represents the number of gray levels, and is the element of GLCM in th row and th column.

The PanTex feature of each pixel is defined as the minimum of the contrast metrics in all GLCMs, and pixels with PanTex feature value larger than a threshold was extracted as buildings.

1. For column (d) and (e), The Morphological Building Index (MBI) of each pixel was calculated by the integration of Differential Morphological Profiles (DMPs) generated by multi-scale and multi-direction White Top-Hat (WTH) by reconstruction transform of the brightness image:

where DMP-WTH represents DMPs generated by WTH operator, and are the scale and direction of each WTH operator which can be adjusted according to the image resolution and building size, and and represent the number of scales and directions. Pixels with larger than a threshold was extracted as buildings.

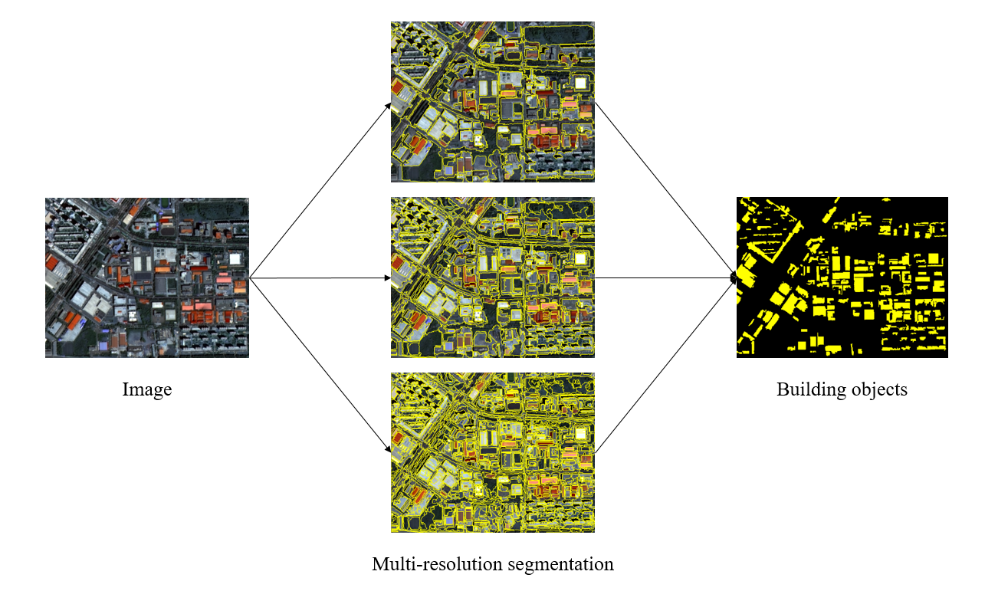
More details for the ideas of these features can be seen in Liu et al. (2019).

Reference:

Liu, C., Huang, X., Zhu, Z., Chen, H., Tang, X., Gong, J., 2019. Automatic extraction of built-up area from ZY3 multi-view satellite imagery: Analysis of 45 global cities. Remote Sensing of Environment, 226: 51-73.

1. Detailed steps involved in the generation of Fig. 5

Firstly, the image was divided into segments using multi-scale mean shift based segmentation (Huang and Zhang, 2008) at the scale of 200, 100 and 50. Then, each building object was identified at the optimal scale among the three different scales by manual interaction. Finally, buildings identified at each scale were integrated as the building extraction result.



**Fig. 5.** An example of mean shift based building detection (Huang and Zhang, 2008). Left: image scene. Middle: three segmentation maps at large, middle, and small scales. Right: the final building detection results.

Reference:

Huang, X., Zhang, L., 2008. An adaptive mean-shift analysis approach for extraction and classification from urban hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing, 46(12): 4173–4185.

1. Specification of Fig. 6

For each dataset, two representative images with buildings of different sizes, shapes and distributions were selected for display. For SpaceNet7, Zeebrugge and Postdam datasets, RGB true-color images were presented. For Vaihingen dataset, images have only NIR, R and G bands, so false-color images were presented. The website to download these datasets are as follows:

<http://www.classic.grss-ieee.org/community/technical-committees/data-fusion/2015-ieee-grss-data-fusion-contest/>

<https://spacenet.ai/sn7-challenge/>

https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-vaihingen/

https://www2.isprs.org/commissions/comm2/wg4/benchmark/2d-sem-label-potsdam/

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| 1. SpaceNet 7 | 1. Zeebrugge | 1. Vaihingen | 1. Potsdam |

**Fig. 6.** Building patches in the open access datasets: (a) two true-color parcels from Paris in the SpaceNet dataset, (b) two true-color parcels in the Zeebrugge dataset, (c) two false-color parcels in the Vaihingen dataset, and (d) a true-color parcel in the Potsdam dataset. These patches can be freely downloaded from the websites listed in Table 1.

1. Detailed steps involved in the generation of Fig. 7

|  |  |  |  |
| --- | --- | --- | --- |
| (a) |  |  |  |
| (b) |  |  |  |
| (c) |  |  |  |
|  | Image | Initial result | Post-processing |

**Fig. 7.** Examples of post-processing for building detection: a) spectral constraint, b) shape verification, and c) shadow verification. Left column: false-color high spatial resolution images; middle column: initial building footprint generated by the method proposed by Huang and Zhang (2011).

1. The spectral constraint contains Normalized Difference Vegetation Index (NDVI) and hue component based correctness improvement to alleviate the false alarms caused by vegetation and soil, the details are as follows:

NDVI was calculated as follows:

The hue component of the image in HSV color space was calculated as follows:

Thresholds were set for and respectively and building pixels were removed from the building map if the following condition was satisfied:

1. Area and length-and-width ratio were used for geometrical constraints. For each connected component in the binary building map, its area and length-and-width ratio were calculated. Components with area smaller than a threshold or length-and-width ratio larger than a threshold were removed from the building map.
2. Shadow was extracted from the image using Morphological Shadow Index (Huang and Zhang, 2013), which was calculated by the integration of DMPs of generated by multi-scale and multi-direction Black Top-Hat (BTH) by reconstruction transform. The shadow map was dilated using a linear structural element, and connected components which were not intersected with the dilated shadow map were removed from the building map.

Reference:

Huang, X., Zhang, L. 2011. A multidirectional and multiscale morphological index for automatic building extraction from mutispectral GeoEye-1 imagery. Photogrammetric Engineering and Remote Sensing, 77(7): 721–732.

Huang, X., Zhang, L. 2013. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5(1): 161-172.

Huang, X., Yuan, W., Li, J., Zhang, L. 2017. A New Building Extraction PostProcessing Framework for High-Spatial-Resolution Remote-Sensing Imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(2): 654-668.

1. Detailed steps involved in the generation of Fig. 8

For incomplete buildings, the holes in the building objects were filled using the imfill() function in MATLAB 2021b. The Pseudo code is:

Post-processing map = [imfill(initial map,'holes')](https://ww2.mathworks.cn/help/images/ref/imfill.html?searchHighlight=imfill&s_tid=srchtitle_imfill_1#d123e24076)

More details can be seen at: <https://ww2.mathworks.cn/help/images/ref/imfill.html?lang=en>

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
| Image | Initial result | Post-processing |

**Fig. 8.** Examples of filling the holes and supplementing the incomplete buildings