

# APPLICATION ORIENTED QUALITY EVALUATION OF GAOFEN-7 OPTICAL STEREO SATELLITE IMAGERY

J. Tian<sup>1,\*</sup>, X. Zhuo<sup>1</sup>, X. Yuan<sup>1</sup>, C. Henry<sup>1</sup>, P. d'Angelo<sup>1</sup>, T. Krauss<sup>1</sup>

<sup>1</sup> Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany  
{jiaojiao.tian, xiangyu.zhuo, xiangtian.yuan, corentin.henry, pablo.angelo, thomas.krauss}@dlr.de

Commission I, WG I/8

**KEY WORDS:** GF-7, Deep Learning, satellite stereo imagery, 3D reconstruction, building segmentation, road segmentation

## ABSTRACT:

GaoFen-7 (GF-7) satellite mission is further expanding the very high resolution 3D mapping application. Carrying the first civilian Chinese sub-meter resolution stereo satellite sensors, GF-7 satellite was launched on November 7, 2019. With 0.65 meter resolution on backward view and 0.8 meter resolution forward view, GF-7 has been designed to meet the demand of natural resource monitoring, land surveying, and other mapping applications in China. The use of GF-7 for 3D city reconstruction is unfortunately restricted by the fixed large stereo view angle of forward and backward cameras with +26 and -5 degrees respectively which is not optimal for dense stereo matching in urban regions. In this paper we intensively evaluate the quality of the GF-7 datasets by performing a series of urban monitoring applications, including road detection, building extraction and 3D reconstruction. In addition, we propose a 3D reconstruction workflow which uses the land cover classification result to refine the stereo matching result. Six sub-urban regions are selected from the available datasets in the middle of Germany. The results show that basic elements in urban scenes like buildings and roads could be detected from GF-7 datasets with high accuracy. With the proposed workflow, a 3D city model with a visually observed good quality can be delivered.

## 1. INTRODUCTION

The 3-dimensional (3D) urban object extraction and reconstruction are among the most essential remote sensing research topics and are demanded for intelligent city management. Satellite stereo imagery is until now the only solution for 3D city monitoring of large regions in high spatial and temporal resolution. Therefore, many countries have started their space program with very high resolution (VHR) stereo satellites (Tian, 2013, Jérôme, 2019, Tang et al., 2020a).

Researches are exploring the potential of VHR satellite stereo imagery since the last century. In 1997 Ridley et al. (Ridley et al., 1997) have used airborne data to simulate the 1 m panchromatic and 4 m resolution multi-spectral stereo data for Digital Surface Model (DSM) generation, change detection and 3D urban modelling, which has proved the potential of using 1 meter resolution satellite data for national wide 3D building modelling. The first civilian very high resolution along track stereo satellite – IKONOS – was launched on 24. September 1999 (Dial et al., 2003), providing a resolution of 0.82 m to 1 m. Limited to the image processing and stereo matching techniques, the initial research using IKONOS data focused on image radiometry and geometric evaluation. Stereo imagery were mainly served for visualization and manually extraction purpose (Dial et al., 2003, Baltsavias et al., 2001, Tao et al., 2004). The automatic DSM generation approaches with different matching techniques were proposed years later (Krauß et al., 2005, Baltsavias et al., 2006, Zhang, Gruen, 2006). The advanced dense matching techniques and the available 0.5 meter resolution WorldView-2 data have brought a new era for the automatic building 3D reconstruction (Tian et al., 2017). Former Digital Globe, now Maxar has further improved the image resolution to 31 cm in their WorldView-3 and Worldview-4 satellite

which were launched in 2014 and 2016, respectively. With the launch of GeoEye in 2009 and Pléiades in 2011, more VHR stereo satellite data are available. However, these VHR stereo satellites only capture stereo data for some specific regions according to demands, and are rather expensive. Therefore, the related researches are still limited to specific small regions ordered as (multi-) stereo data. The specially designed stereo satellites Cartosat-1 and ALOS/PRISM only provide 2.5 meter resolution data. Cartosat-1 captures just panchromatic data, and ALOS/PRISM is no more in operational mode. Therefore, more stereo data are still demanded for a global 3D monitoring.

China has launched the first high resolution stereo satellite Ziyuan-3 (ZY3) on 9 January 2012, with 2.1 meter resolution in nadir view. The forward and backward camera that are inclined at  $\pm 22^\circ$  can provide stereo images with a resolution of 3.5 meter (Tang et al., 2020a). GF-7 is the first civilian sub-meter stereo satellite in China. Different to the Ziyuan-3 series, GF-7 system is composed of two cameras with forward and backward views, respectively (Tang et al., 2020b).

Several applications and researches on GF-7 dataset are available. However, most of them are concentrating on the Laser Altimeter System (Tang et al., 2020b), positioning accuracy (Liu et al., 2021). Although GF-7 stereo imagery have been available since two years ago, the utilization of these data for 3D reconstruction is restricted partly due to the large stereo view angle, which brings extra difficulty in stereo matching over urban region. In the present paper, we examine the quality of nadir view very high resolution multi-spectral data in the application of building detection and road detection with state-of-the-art deep neural networks, as well as the building 3D reconstruction results after the object-based refinement. A workflow has been specially designed for GF-7 stereo imagery accordingly.

\* Corresponding author, all authors have contributed equally



Figure 1. Pan-sharpened images of the 6 selected test regions, each  $4000 \times 4000 \text{ m}^2$

## 2. GF-7 DATASET

### 2.1 GF-7 mission

GF-7 has backward view cameras viewing a near nadir view direction. It can provide backward panchromatic images with 0.65 meter resolution and four bands multi-spectral images with 2.6 meter resolution. The forward panchromatic image with  $26^\circ$  inclination has a resolution of 0.8 meter (GF-7 Satellite, n.d.). Detailed parameters of the GF-7 two-line stereo imagery are shown in Table.1. As another highlight, GF-7 is equipped with a laser altimeter system with  $1.6 \text{ km} \times 1.6 \text{ km}$  plot size. But these data are not involved in this manuscript.

Item	Parameters	
Spectral Bands	Pan	0.45 ~ 0.90 $\mu\text{m}$
		0.45 ~ 0.52 $\mu\text{m}$
	MS	0.52 ~ 0.59 $\mu\text{m}$
		0.63 ~ 0.69 $\mu\text{m}$
		0.77 ~ 0.89 $\mu\text{m}$
Forward inclination	$+26^\circ$	
Backward inclination	$-5^\circ$	
Resolution	Pan	Backward Pan: 0.65 m Forward Pan: 0.8 m
	MS	Backward MS: 2.6 m
Swath Width	20 km	

Table 1. Description of GF-7 two-line stereo imagery (GF-7 Satellite, n.d.).

### 2.2 Test regions

The test regions are located in Hesse, Germany, to the south west of Frankfurt. Two scenes were captured on 21. July 2020 and provided for our experiment. We have selected six urban/sub-urban regions with dense building distributions from them. Each test region has a size of  $4000 \times 4000 \text{ m}^2$  as shown in Figure 1.

### 2.3 Data processing

We have followed a standard workflow for DSM generation and orthoimages preparation without ground control points. A coarse to fine matching procedure which has been proposed for stereo satellite ZY-3 and Cartosat-1 data is adopted for GF-7 data (d'Angelo, 2013). To further improve the absolute accuracy of the Rational Polynomial Coefficients (RPCs) from the GF-7 data, Shuttle Radar Topography Mission (SRTM) with a resolution of 30 meter and geo-accuracy below 10 meter is used as absolute geolocation references. This approach works well over the selected regions as they are characterized by hilly terrains. Afterwards, multi-ray tie points are automatically measured between the forward and backward panchromatic images using pyramidal local least square matching. A bundle block adjustment based on these tie points is performed (Grodecki, Dial, 2003). After the Rational Polynomial Coefficient (RPC) correction, DSMs are generated with Semi-Global matching, which is still the most robust dense matching approach by considering both efficiency and accuracy (Xia et al., 2020). In this step, we use the Census cost function as similarity measure. The resulting DSM has 1 meter resolution, which is approximately two times of the original image resolution. In the end,

the delta surface fill algorithm (Grohman et al., 2006) is performed to fill the unmatched pixels with SRTM heights and generate the final DSM. It has to be noted that the whole DSM generation procedure is fully automatic without any manual processing.

Orthophotos of the panchromatic and multispectral images are generated with the filled DSM and refined RPC parameters. As backward view cameras of GF-7 have nearly nadir view, the quality of the orthophotos is benefiting from it. To fully exploit the spatial resolution of GF-7, we carry out a Gram-Schmidt pansharpening with ENVI software on the backward panchromatic and multispectral data (Laben, Brower, 2000). The pansharpened images of 6 test regions are shown in Figure.1.

### 3. BUILDING FOOTPRINT EXTRACTION

#### 3.1 Building segmentation

Building segmentation is one of the most crucial applications of VHR satellite imagery. With image resolution or Ground Sampling Distance (GSD) sufficient to exhibit ground objects such as buildings and roads, it could provide accurate wide-range coverage over the earth's surface. Building footprints are indispensable for urban planning, disaster relief, map services etc. With the rise of deep learning based approaches, building footprints can be extracted more accurately and efficiently for use in innumerable disciplines. To evaluate the quality of GF-7 satellite imagery, we carry out building footprint extraction using a semantic segmentation deep learning model that was trained with a public benchmark dataset.

We used an off-the-shelf neural network named High-Resolution Network (HRNet) (Sun et al., 2019) for the building segmentation, which maintains high-resolution representations of the image through the network. Please refer to the original paper for more details about the model. Currently no public building segmentation dataset employs GF-7 imagery. We use satellite images from the xBD dataset with similar spatial resolution as training data (Gupta et al., 2019). Provided by the Maxar Open Data Program, xBD dataset is originally created for building damage classification. Specifically, the dataset comprises of coregistered pre- and post-disaster VHS satellite images. To train the building segmentation model, building boundary ground truth is converted to building masks. The ground sampling distance (GSD) is around 46 cm. The xBD dataset has a geographical coverage including North America, Southeast Asia, Portugal and Australia, encompassing various building and landcover types.

We employ a simple pre-processing for GF-7 imagery in our experiment. The 2 % and 98 % quantities of the images are stretched to values between 0 and 255. In both training and testing, images are tiled to patches of size  $1024 \times 1024$  pixels<sup>2</sup>. In testing, the output is a softmax probability map, and 50 % overlap is used when stitching. To eliminate the boundary effect, we use a square weight matrix which weights down prediction probability of pixels closer to the boundary.

#### 3.2 Building footprint vectorization

Typically, state of the art building extraction methods generate pixel-wise building segmentation. However, these building segments have to be converted into vector formats before they can be directly used by the mapping agencies (Girard et al.,

2021). Manual identification and delineation of buildings from VHR remote sensing imagery is extremely time-consuming and unrealistic for large-scale datasets. Therefore we proposed an automated footprint vectorization strategy to further refine the building segmentation results. First, we extract initial building corner points from the segmentation results; second, we recover critical corner points that are wrongly removed in the first step; finally, we adjust the position of these building corners via a geometry-based optimization.

**3.2.1 Polygon Initialization** Although deep learning based segmentation approaches can generally achieve good segmentation results in terms of standard accuracy evaluation metrics, they suffer from the limited localization ability of Convolutional Neural Networks (CNNs) and often result in blob-like segments, smooth corners and inaccurate object boundaries. In this case, corner detection algorithms like the Harris and Förstner detectors may remove many critical vertices. In order to ensure the existence of potential building corners, we first extract all the pixels on the boundary as candidates, and then apply the Douglas-Peucker algorithm (Douglas, Peucker, 1973) to filter out co-linear points. Afterwards, over-short edges, over-sharp or over-smooth corners are further removed by giving pre-defined thresholds. The remaining vertices are accepted as initial building corners.

**3.2.2 Corner points recovery** With a rigid threshold, the Douglas-Peucker algorithm may often remove critical corners due to the "zig-zag" pattern of boundary lines. Considering the fact that non-adjacent edges of a normal building are mostly parallel, we assume that the edges have two or more dominant tangent directions. In order to recover these critical corners, we first detect the dominant tangent directions of the polygon by voting the directions of all edges. Those edges that deviate far from all the tangent directions are considered to have missing corners, and a corner point will be added in between.

**3.2.3 Geometry-based polygon optimization** The positions of the corner points are further adjusted via optimization based on geometry-based rules. The energy function is composed by following terms:

- Alignment between the vectorized polygon and the building segmentation mask. The vectorized polygon is encouraged to be aligned with the original building mask, and the deviation is measured by their chamfer distance (Borgefors, 1986).

$$L_{poly\_mask} = \frac{1}{N} \sqrt{\sum_{l \in L_{poly}} d_{I(l)}^2}, \quad (1)$$

where  $L_{poly}$  denotes the edge model of the polygon,  $d_{I(l)}$  stands for the distance values where the edge model  $L_{poly}$  hits the distance image calculated from the building mask, whereas  $N$  is the number of points in  $L_{poly}$ .

- Orthogonality of the vectorized polygon. The vectorized polygon is enforced to have regularized shapes, i.e., adjacent edges should be perpendicular with each other.

$$L_{ortho} = \frac{1}{N} \sqrt{\sum_i^N (\theta_i - 90)^2} \quad (2)$$

Where  $\theta_i$  denotes the tangent angle of an edge  $i$  and  $N$  is the number of edges in a polygon.

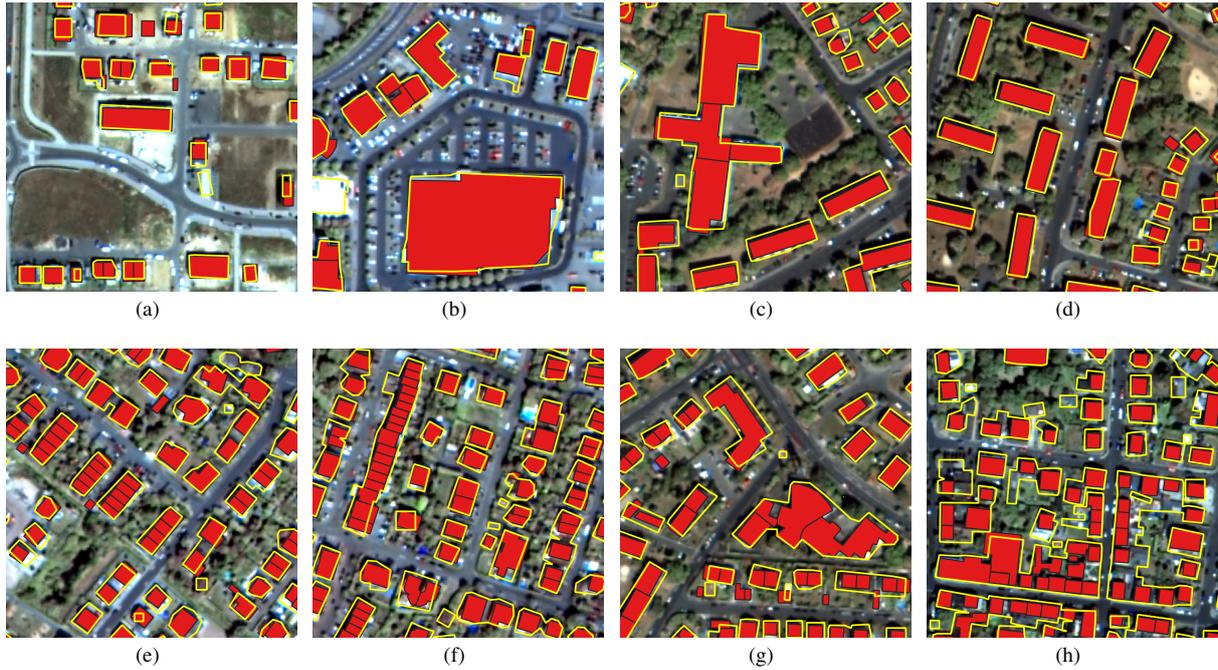


Figure 2. Building extraction results in yellow and the reference masks (red) overlaid on orthophotos.

RoI	Segmentation		Vectorization	
	F1 Score[%]	IoU [%]	F1 Score[%]	IoU [%]
1	75.96	61.23	74.29	60.59
2	74.57	59.45	74.62	60.23
3	78.83	65.06	76.98	64.62
4	73.13	57.64	73.92	55.92
5	76.47	61.90	77.02	62.20
6	75.54	60.69	74.13	60.34

Table 2. F1 score and IoU results of building segmentation in GF-7 satellite imagery.

- Consistency of tangent directions. Strict constraints of regularization may lead to zig-zac effect of the polygon, therefore we also encourage edges to align with at least one of dominant tangent directions of the polygon.

$$L_{tangent} = \frac{1}{N} \sqrt{\sum_i^N \min_{\theta_j \in \theta_D} ((\theta_i - \theta_j)^2)} \quad (3)$$

Where  $\theta_i$  denotes the tangent angle of an edge  $i$ ,  $\theta_j$  the tangent value of a dominant direction that belongs to the set of all dominant directions  $\theta_D$ , and  $N$  is the number of edges in a polygon.

The total loss is a linear combination of the above losses with individual coefficients:

$$L_{total} = L_{poly-mask} + \lambda_1 L_{ortho} + \lambda_2 L_{tangent}, \quad (4)$$

The hyperparameters  $\lambda_1$  and  $\lambda_2$  are manually tuned. In experiment we set  $\lambda_1$  0.2 and  $\lambda_2$  0.5.

### 3.3 Evaluation

For evaluating the accuracy of the extracted building footprint, we have corrected the building footprints from open street map (OSM) to generate the reference datasets. We have refined the co-registration between the OSM maps and our orthophotos and manually edited the building polygons with conspicuous errors. The reference masks match well with the images from both building locations and boundaries. As shown in Figure 2, we have overlaid the extracted building footprint and reference mask to the orthophoto. The building reference mask are shown as red masks. Our extracted building boundaries are displayed as yellow polygons. As can be seen, the prepared reference masks have a generally high quality and match very well with the orthophotos. Our approach could precisely extract the building boundaries for the buildings with regular shapes and large size (Figure. 2 a-d). For the rectangular shaped buildings, our footprint vectors match precisely to the OSM footprints with sharp boundaries and right angles. After removed the 'zig-zag' pattern of boundaries, our approach would still preserve the main shape of the buildings with our corner points recovery approach. However, for very dense distributed buildings, two or more buildings will be recognised as one building (Figure. 2 e-h).

In addition, we calculate the F1 score and Jaccard index (intersection over union, IoU) for the building masks generated from the segmentation and vectorization step, respectively. The results are summarized in Table. 2. It can be seen that the vectorized footprints have comparable or even lower semantic accuracy, this is because the improvement of regularity is at the cost of semantic accuracy.

## 4. ROAD EXTRACTION

### 4.1 Road Segmentation

In the same way as buildings, extracting roads in remote sensing imagery is critical for following the expansion of cities and

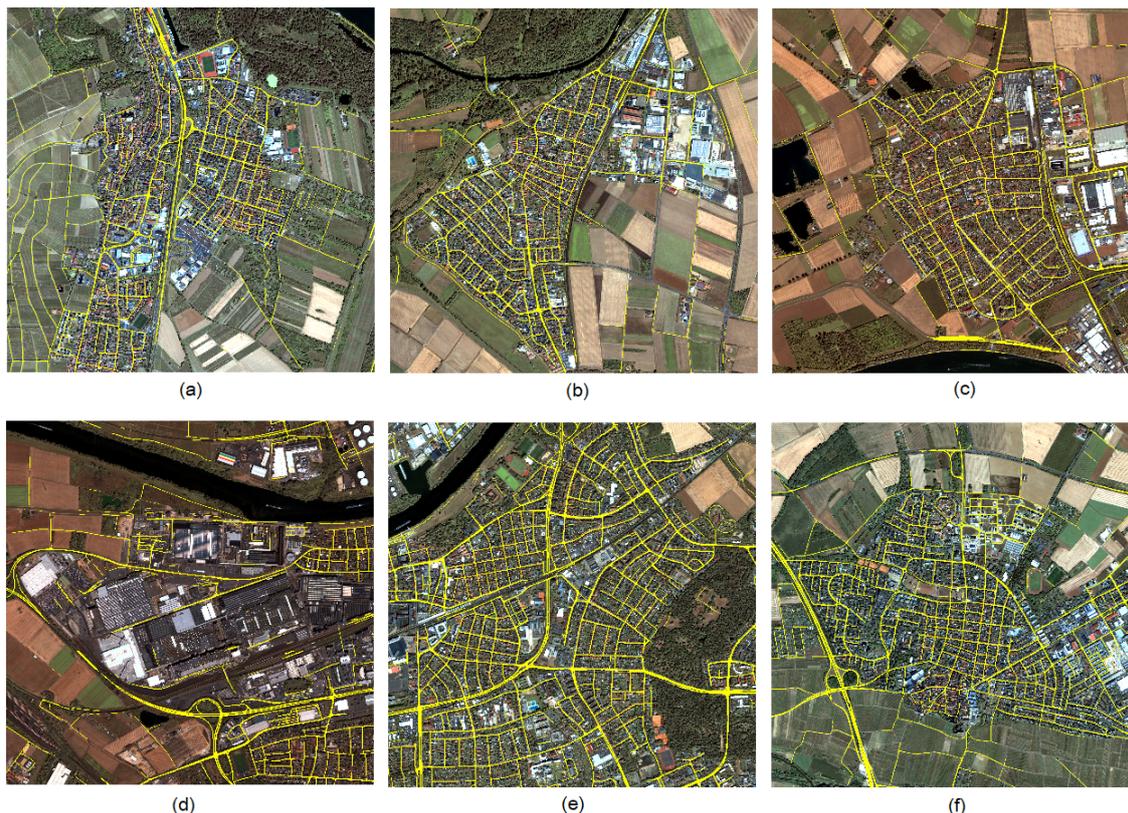


Figure 3. Detected road masks of the six test regions

surveying the availability of transportation networks throughout a country. Roads are however notably difficult to locate and segment properly due to their usually thin outline and the effect of occlusion from other higher objects like vegetation and buildings (Demir et al., 2018, Henry et al., 2021b). The most effective methods for tackling these challenges are fully-convolutional neural networks (FCNs) (Long et al., 2015) and especially models derived from the widely adopted U-Net architecture: such methods have been successfully applied to datasets like the Massachusetts Roads Datasets (Mnih, 2013, Mosinska et al., 2018, Henry et al., 2021a), the 2018 DeepGlobe Road Extraction Challenge (Demir et al., 2018, Zhou et al., 2018) and the SpaceNet 3 Roads Extraction and Routing Challenge (MaxarTechnologies, 2018, Buslaev et al., 2018). We use a Dense-U-Net-121 (Henry et al., 2021a), which is composed of a Dense-Net-121 (Huang et al., 2017) encoder and a decoder based on a mirrored Dense-Net-121.

#### 4.2 Evaluation

Our model is trained on around 5000 images from the DeepGlobe dataset (Demir et al., 2018), i.e. on 50 cm/px GSD imagery from south-eastern Asian regions. Although the images from the GF-7 dataset are acquired at 60 cm/px, the ground resolution difference is small enough not to cause a significant performance loss. We use same image patches as prepared in the building segmentation step. To reduce the visibility of seams between patches when stitching the results together, the road probabilities are predicted, merged with pixel-wise max-voting on the overlapping regions, and only then thresholded into 0s and 1s.

The extracted road masks of all six test regions are shown in

Figure 3. These regions feature various types of roads, from asphalted streets to country roads as well as highways, with different widths and surrounded by a variety of contexts like residential buildings, commercial and industrial areas, fields and forests. Despite the challenges posed by these widely different scenarios, the model demonstrated an equally good performance on each of them. The extracted roads match the real-world topology in, width, smoothness and connectivity across all images. A confusion with railways is visible in image Figure 3 (d) where a few road segments were wrongly predicted, but this is a known issue of such model. Models trained on DeepGlobe indeed generalize well to other images and scenarios, but are still limited to the scenarios covered in the training set. Our model nonetheless performed well out-of-the-box, as anticipated, showing that the GF-7 dataset’s images are similar enough and of high enough quality to be used as a benchmark for training or testing road extraction once annotations become available. Alternatively, readily available ground truth could also be used if co-registrations offsets are small or can be mitigated.

### 5. 3D RECONSTRUCTION

Though GF-7 has a sub-meter resolution and the ability to obtain stereo view imagery, the large stereo view angle has largely restricted the quality of the generated DSMs, especially in urban regions. To refine these DSMs, we firstly generate the normalized DSM (nDSM) and digital terrain model (DTM) using morphological top-hat reconstruction as mentioned in (Qin et al., 2016). Afterward, we project the extracted building footprint vectors to the corresponding normalized DSM (nDSM), with which we generate the a max building height image. So

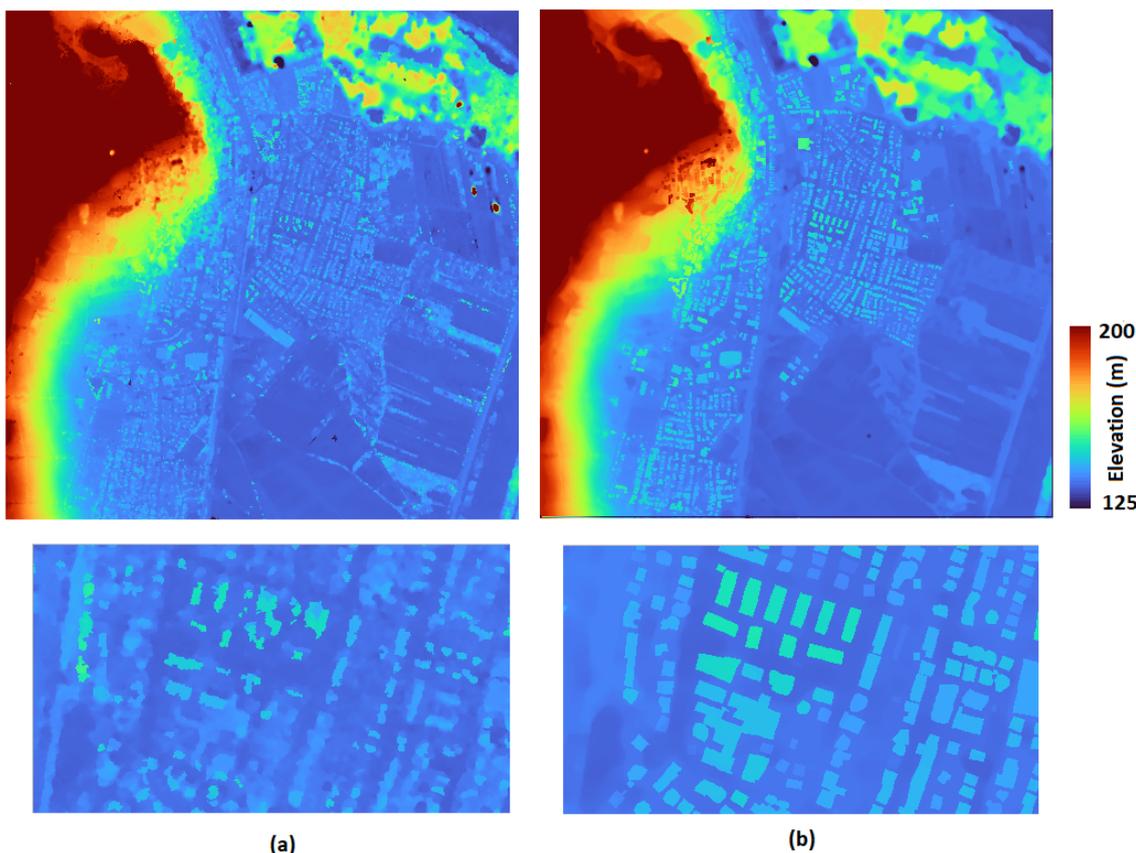


Figure 4. Comparison of the original generated DSM (a) and refined DSM (b).



Figure 5. 3D view of the refined DSM model textured by the orthophoto and road segmentation.

that every pixel inside a single building block receives the max value of the corresponding building height regarding the refined heights, which will be added to the DTM and deliver the final refined DSM. In Figure. 4, the originally generated DSM and the refined DSM of test region-1 are shown in (a) and (b), respectively. One can easily compare visually the quality of these two DSMs. Especially in the refined DSM, most of the buildings have correctly received a reasonable elevations that are 5 to 15 meters higher than the elevation of roads. We have only used the spectral images for building extraction and vectorization due to the limited quality of the initial DSM (Figure. 4 (a)).

A 3D city model view is visualized in Figure. 5 by texturing the refined DSM with corresponding orthophoto and road segments.

## 6. CONCLUSION

In this paper, we evaluate the quality of the GF-7 stereo datasets by applying of the state-of-art building footprint extraction, road segmentation and 3D reconstruction approaches. With 0.65 meter resolution, the pansharpened orthophotos have comparable quality as other sub-meter resolution satellite data. Thus, the building and road segmentation models which are pretrained using respectively xdb and DeepGlobe benchmark datasets perform well on GF-7 datasets. The vectorization step proposed in this paper has further improved the building masks and results in much sharper and more precise building footprints in vector format. To improve the quality of the DSMs that are derived from stereo matching, we presume that all pixels covered by a single roof should have a identical elevation height, thus use the building footprint map to refine the generated DSM. Gable roof and complicated roof shapes are not considered in our current work. The results are promising, a primary 3D city model can be generated from our first results.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge Land Satellite Remote Sensing Application Center, Ministry of Natural Resources of the People's Republic of China for providing the GF-7 datasets.

## REFERENCES

Baltsavias, E., Li, Z., Eisenbeiss, H., 2006. DSM generation and interior orientation determination of IKONOS images using a testfield in Switzerland. *Photogrammetrie Fernerkundung Geoinformation*, 2006(1), 41.

- Baltsavias, E. P., Pateraki, M. N., Zhang, L., 2001. Radiometric and geometric evaluation of ikonos geo images and their use for 3d building modelling. *Joint Workshop of ISPRS Working Groups I/2, I/5 and IV/7 High Resolution Mapping from Space 2001*, ETH Hönggerberg, Institute of Geodesy and Photogrammetry.
- Borgefors, G., 1986. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing*, 34(3), 344 - 371.
- Buslaev, A., Seferbekov, S., Iglovikov, V., Shvets, A., 2018. Fully convolutional network for automatic road extraction from satellite imagery. *CVPR Workshops*.
- Demir, I. et al., 2018. DeepGlobe 2018: A Challenge to Parse the Earth Through Satellite Images. *CVPR Workshops*.
- Dial, G., Bowen, H., Gerlach, F., Grodecki, J., Oleszczuk, R., 2003. IKONOS satellite, imagery, and products. *Remote sensing of Environment*, 88(1-2), 23–36.
- Douglas, D. H., Peucker, T. K., 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2), 112–122.
- d'Angelo, P., 2013. Evaluation of ZY-3 for DSM and ortho image generation. *International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, 1, 57–61.
- GF-7 Satellite, n.d. <http://www.sasclouds.com/english/satellite/chinese/gf7>. Accessed: 2021-08-18.
- Girard, N., Smirnov, D., Solomon, J., Tarabalka, Y., 2021. Polygonal building extraction by frame field learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5891–5900.
- Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering & Remote Sensing*, 69(1), 59–68.
- Grohman, G., Kroenung, G., Strebeck, J., 2006. Filling SRTM voids: The delta surface fill method. *Photogrammetric Engineering and Remote Sensing*, 72(3), 213–216.
- Gupta, R., Goodman, B., Patel, N., Hosfelt, R., Sajeev, S., Heim, E., Doshi, J., Lucas, K., Choset, H., Gaston, M., 2019. Creating xbd: A dataset for assessing building damage from satellite imagery. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 10–17.
- Henry, C., Fraundorfer, F., Vig, E., 2021a. Aerial Road Segmentation in the Presence of Topological Label Noise. *2020 25th International Conference on Pattern Recognition (ICPR)*, 2336-2343.
- Henry, C., Hellekes, J., Merkle, N., Azimi, S., Kurz, F., 2021b. Citywide estimation of parking space using aerial imagery and osm data fusion with deep learning and fine-grained annotation. *ISPRS 2021*, XLIII, 479–485.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely Connected Convolutional Networks. *CVPR*.
- Jérôme, S., 2019. Shaping the future of earth observation with pléiades neo. *2019 9th International Conference on Recent Advances in Space Technologies (RAST)*, IEEE, 399–401.
- Krauß, T., Reinartz, P., Lehner, M., Schroeder, M., Stilla, U., 2005. DEM Generation from Very High Resolution Stereo Satellite Data in Urban Areas Using Dynamic Programming. *ISPRS Workshop, Hannover*, 36 (1/W3).
- Laben, C. A., Brower, B. V., 2000. Process for enhancing the spatial resolution of multispectral imagery using panch sharpening. US Patent 6,011,875.
- Liu, Y., Mo, F., Zhang, Y., Xie, J., Li, Q., Hu, F., Liu, C., 2021. Validation of preliminary geometric positioning accuracy for China's civil high-resolution surveying and mapping satellites: Ziyuan-3 01 and Gaofen-7 panchromatic imagery. *Remote Sensing Letters*, 12(6), 521–530.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully Convolutional Networks for Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440.
- MaxarTechnologies, 2018. SpaceNet on Amazon Web Services (AWS). Datasets. The SpaceNet Catalog.
- Mnih, V., 2013. Machine Learning for Aerial Image Labeling. PhD thesis, University of Toronto.
- Mosinska, A., Marquez-Neila, P., Kozinski, M., Fua, P., 2018. Beyond the Pixel-Wise Loss for Topology-Aware Delineation. *CVPR*.
- Qin, R., Tian, J., Reinartz, P., 2016. Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images. *International Journal of Remote Sensing*, 37(15), 3455–3476.
- Ridley, H. M., Atkinson, P. M., Aplin, P., Muller, J.-P., Dowman, I., 1997. Evaluating the potential of the forthcoming commercial US high-resolution satellite sensor imagery at the ordnance survey (R). *Photogrammetric engineering and remote sensing*, 63(8), 997–1005.
- Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5693–5703.
- Tang, X., Gao, X., Cao, H., Mo, F., Wang, Z., Xu, W., Zhu, G., Yue, Q., Hu, F., Zhu, H. et al., 2020a. The China ZY3-03 Mission: surveying and mapping technology for high-resolution remote sensing satellites. *IEEE Geoscience and Remote Sensing Magazine*, 8(3), 8–17.
- Tang, X., Xie, J., Liu, R., Huang, G., Zhao, C., Zhen, Y., Tang, H., Dou, X., 2020b. Overview of the GF-7 laser altimeter system mission. *Earth and Space Science*, 7(1), e2019EA000777.
- Tao, C. V., Hu, Y., Jiang, W., 2004. Photogrammetric exploitation of IKONOS imagery for mapping applications. *International Journal of Remote Sensing*, 25(14), 2833–2853.
- Tian, J., 2013. 3D change detection from high and very high resolution satellite stereo imagery. PhD thesis, Universität Os-nabrück.

Tian, J., Metzloff, L., d'Angelo, P., Reinartz, P., 2017. REGION-BASED BUILDING ROOFTOP EXTRACTION AND CHANGE DETECTION. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42.

Xia, Y., d'Angelo, P., Tian, J., Reinartz, P., 2020. Dense matching comparison between classical and deep learning based algorithms for remote sensing data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43(B2), 521–525.

Zhang, L., Gruen, A., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(3), 195–211.

Zhou, L., Zhang, C., Wu, M., 2018. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. *CVPR Workshops*.