Potential of Satellite Image in the End-To-End 3D City Modelling for Developing Regions: A Case Study for Cau Giay District, Hanoi, Vietnam

Phan, A.,[1,2*] Bui, H.[1] and Nguyen, T.[1]
[1]Center of Multidisciplinary Integrated Technologies for Field Monitoring (FIMO), University of Engineering and Technology, Vietnam National University Hanoi, 144 Xuanthuy, Hanoi 100000, Vietnam
 E-mail: anhp@fimo.edu.vn
[2]International Digital Earth Applied Science Research Center, Chubu Institute for Advanced Studies, Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi 487-8501, Japan
*Correspondence Author

Abstract

Recently, the problem of missing building foundation and height data is making it difficult to develop 3D GIS in developing countries due to the high cost of accurate data collection and data sharing policy. In this study, we proposed an approach for 3D city modeling that is suitable for these areas, and then the experiment was carried out for Cau Giay district, Vietnam. First, building footprints will be obtained from the free high-resolution satellite imagery source. Then, the building footprints themself and the surrounding area images are used to extract key points by some well-known descriptors. K-means will then be applied to create bag of word (BoW) features, which are then used to build a regression model to estimate the building height over the study area. We obtained the error MAE = 3.5m and RMSE = 33.6 for our height regression model. The results achieved show the ability to apply satellite images to extract fundamental building topologies such as building boundary and building height to initially build a LoD1 3D city model.

1. Introduction

1.1 3D Modelling in the Developed and Developing Countries

3D city is becoming more and more popular in all the countries around the world, especially in the developed countries and cities (e.g. New York City, Berlin, and Helsinki). These 3D geo-information systems have successfully conceptualized the modern cities as level-of-detail-2 (LoD2) 3D building models which contain textured ground surfaces, wall surfaces, and roof surfaces. To build a 3D city in an efficient way, it is necessary to acquire essential building information including building footprint, wall height, roof height, and textures of the building. The lack of each field data or low accuracy makes a huge impact and deficiency in 3D visualization compared to the real-life objects.

The availability of the required data depends mostly on the used technology and the data policy of a country. In particular, to obtain the building footprint and building height data, ones choose to use the administrative database while the others try to collect these from scratch. The first approach has been mainly applied in the developed countries and cities due to the richness of available geo-information in national archive. The second method varies considerably in different countries based on the current development situation of technology. In collecting the required input information (building footprint, building height) process, the most common way relates to the field of computer vision. In particular, satellite images and normal optical images have been used to deploy traditional machine learning and deep learning models to classify and segment the building footprint, to estimate the building height.

Building footprint extraction has been a well-known problem since many competitions were held to devise the solution, thus a huge amount of free data was provided for examination. For example, SpaceNet is the most massive dataset freely provided and hosted as an Amazon Web Services (AWS) Public Dataset, which contains over 11 million building footprints (https://spacenet.ai/datasets/). With a rapid growing amount of data sources including high resolution satellite image, high resolution aerial optical image, and aerial LiDAR data, automatic data processing methods need to be considered with a very particular attention. First, simple thresholds approaches have been widely used in the past to discriminate buildings, cars, trees using a LiDAR Digital Elevation Model (DEM) (Haithcoat et al., 2001).

In another study, Haithcoat et al. proposed an approach to distinguish building from the other objects based on Digital Surface Model (DSM) and geometric characteristics (size, shape and height) (Haithcoat et al., 2001). Also from the aerial LiDAR data, Wang et al., (2006) used a Bayesian approach for building footprint extraction. Neural Networks classifier model was designed by Bittner to extract building footprint from the data combination of normalized Digital Surface Model (nDSM) and Normalized Difference Vegetation Index (NDVI) (Davydova, et al., 2016). In the more recent studies, very high-resolution images from commercial satellites have been commonly used in spite of high cost to improve the performance of building boundary classification and extraction using a range of methods from contour features (Liasis and Stavrou, 2016a) to traditional machine learning and state of the art deep learning models (Zhao et al., 2018 and Li et al., 2019).

Recently, building height estimation has been realized in many study areas as its significant role in 3D city modelling, urban planning and navigation (Zhao et al., 2019). Building height has mostly been estimated from its building shadow captured by high-resolution satellite images and aerial images (Irvin and McKeown, 1989, Liasis and Stavrou, 2016b, Kadhim and Mourshed, 2018 and Qi et al., 2016). Besides, the it was also obtained from other building features such as building footprint combined other attributes available (e.g. storey, age of building, usage, …) from volunteered geoinformation and cadastre with traditional machine learning models and the result achieved a promising mean absolute error (MAE) of 0.8m in the inferred heights by a study conducted in the city of Rotterdam (Biljecki et al., 2017). With a rising application of deep learning, several studies have applied it to tackle the problem of building height estimation from single monocular imagery via a fully residual convolutional-deconvolutional neural network named IM2HEIGHT (Mou and Zhu, 2018) and from the complex street scene images via a deep neural network named BuildingNet (Zhao et al., 2019). In addition, building height and footprint data were also collected and extracted from well-known commercial software such as Pix4D, Global Mapper, and Agisoft Photoscan.

Vietnam is a developing country in Southeast Asia. Although 3D City in Vietnam is no longer a new concept, there are still not many outstanding studies and specific applications of 3D city. In addition, the lack of house foundation and height data is also one of the major problems that hinders the construction of a complete 3D city with high accuracy and practical use. A prominent study has conducted to build a 3DGIS systems (3DVNU) for the Vietnam National University in Hanoi (Anh et al., 2017). Building models are represented at a level of detail LoD2 with textured wall and roof surfaces. Real-time air quality parameters and weather effects are synchronized and shown in the 3D VNU system. Besides, to overcome the lack of height data, Anh et al., (2019) developed a height regression machine learning model for buildings in Hanoi, next the LoD1 3D models of Hanoi were obtained by extruding the building footprint to its height. Another study was carried out by the authors to construct house foundation label data set from high resolution satellite imagery provided by Google Earth (Nguyen et al., 2019). This data set was then used to experimentally build a deep learning model of building footprint classification in Cau Giay, Hanoi.

Through some prominent studies above, it can be seen that around the world, in developed countries and cities, 3D cities have been built and developed. Because there are still many difficulties in data collection, the level of perfection of building models is still mainly limited to LoD2 detail. In Vietnam and some other developing countries, data constraints due to costly collection and usage policies have made it difficult to build 3D cities with LoD2 on large scale. However, a number of studies on a small scale have contributed to proposing possible solutions in the development of large 3D city systems in the future.

1.2 Objective of the Study and Outline of the Paper
Based on the issues outlined above, in this study, we proposed and experimented the method of building 3D models with level of detail LoD1 using only free satellite imagery. The goal of the study is to propose a method to solve the basic data shortage problem to build 3D models in developing areas. First, building foundation data in the study area will be collected, this footprint data is then used to mask out satellite images into separate houses. Next, the building height will be estimated based on building footprint and surrounding area images. Finally, 3D city models with level of detail LoD1 on Cau Giay, Hanoi will be produced from extracted building boundary and estimated building height.

2. Materials and Methods
In this section, the introduction of the study area will be presented in section 2.1, the data used in the study are described in section 2.2, the height data regression method is introduced in section 2.3, and finally, the method of building 3D models from height data is presented in section 2.4. An overview of the proposed method is shown in Figure 1.
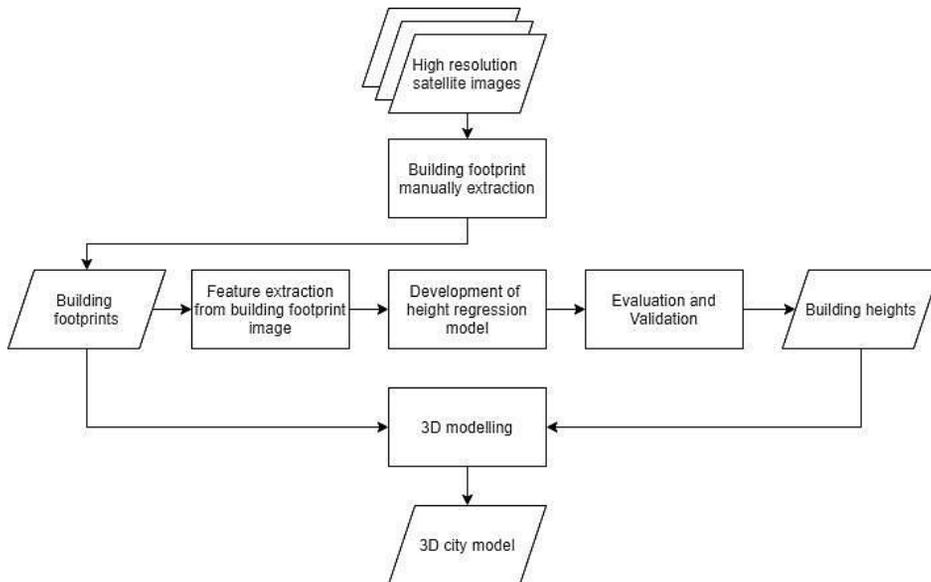
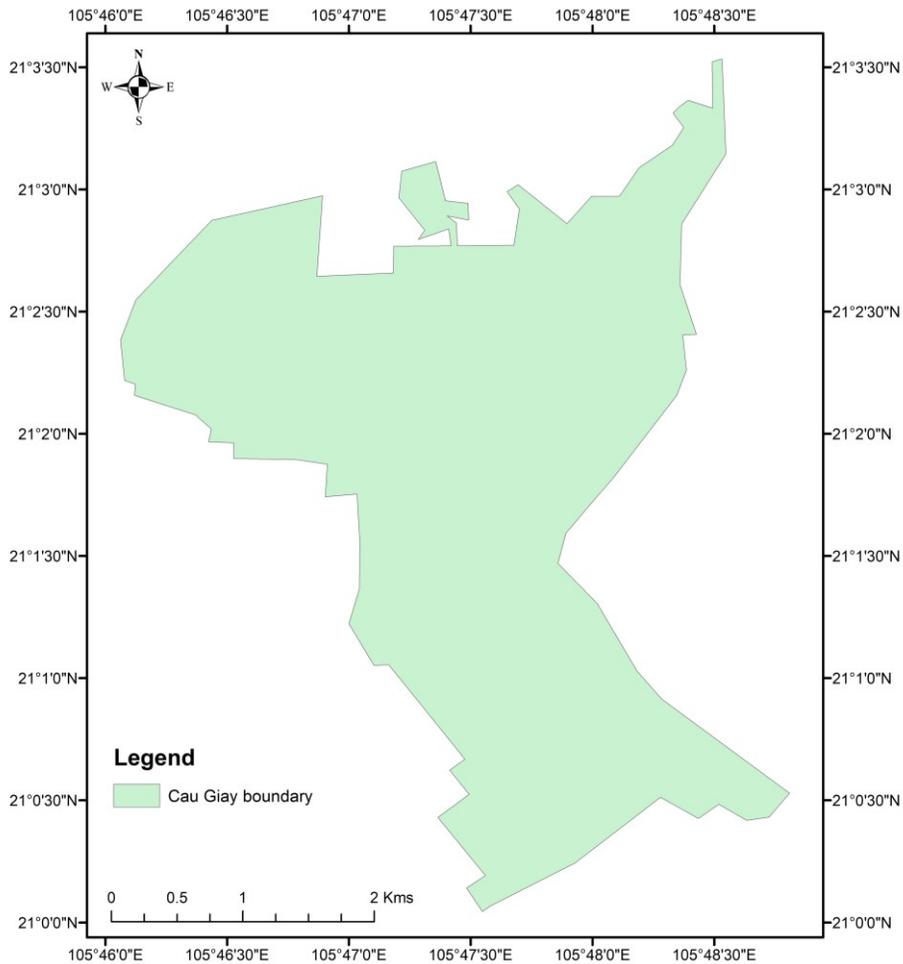Figure 1: Overall flowchart for the proposed methodology

Figure 2: The study area – Cau Giay, Hanoi, Vietnam

First, house foundation data are manually extracted from high-resolution satellite image in the study area. The foundation data set will then be used to crop the images of the building. The building footprint images are then masked to extract features based on the key point detection approach. After the features are extracted, the height-estimated regression machine learning model is constructed and trained. The trained, evaluated and verified model will be used to estimate the height of all buildings in the study area. Finally, house foundation data and height data will be combined to build 3D models for Cau Giay area.

2.1 Study Area

The study area selected for this study is the Cay Giay area. Cau Giay is an inner district of Hanoi city. Cau Giay District is located in the west of Hanoi city center and has 8 affiliated communal administrative units, including 8 wards: Dich Vong, Dich Vong Hau, Mai Dich, Nghia Do, Nghia Tan, Quan Hoa, Trung. Hoa, Yen Hoa. Cau Giay district area is 12.04 km2 with a population of about 269,637 people. The geographic location image of Cau Giay district is shown in Figure 2.

2.2 Data
2.2.1 Satellite images

The satellite image data used in the study is provided free of charge from Google Earth, the satellite image data used is DigitalGlobe Quickbird with a resolution of 65cm pan-sharpened (65 cm panchromatic at nadir, 2.62 m multispectral at nadir). The data is automatically downloaded through the Google Satellite Map Downloader software. The research team has divided the study area into 16 smaller images to facilitate downloading data due to the application's data request limitations. Data provided through Google Satellite Map Downloader software will be divided into small photo scenes (multiple parts of a building), then the subscenes are merged into larger image scenes containing the buildings. Below is an image of a small photo shot of 16 shots in the Cau Giay area (see Figure 3).

2.2.2 Ground truth height data

The ground truth data set was collected based on manual field trip and measurements of buildings and provided by the Vietnam Department of Surveying and Mapping.
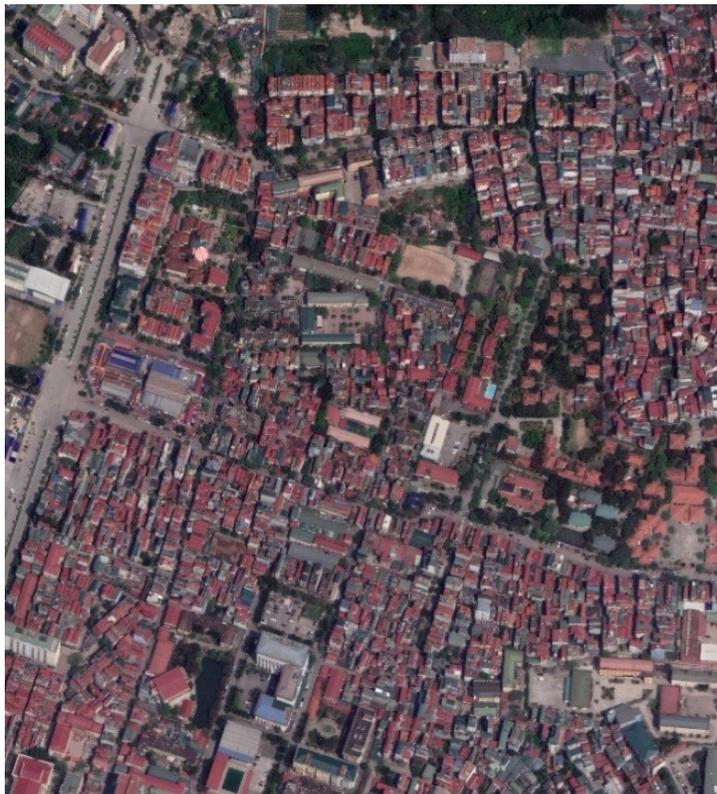

Figure 3: One scene in 16 scenes across the entire study area - Cau Giay, Hanoi
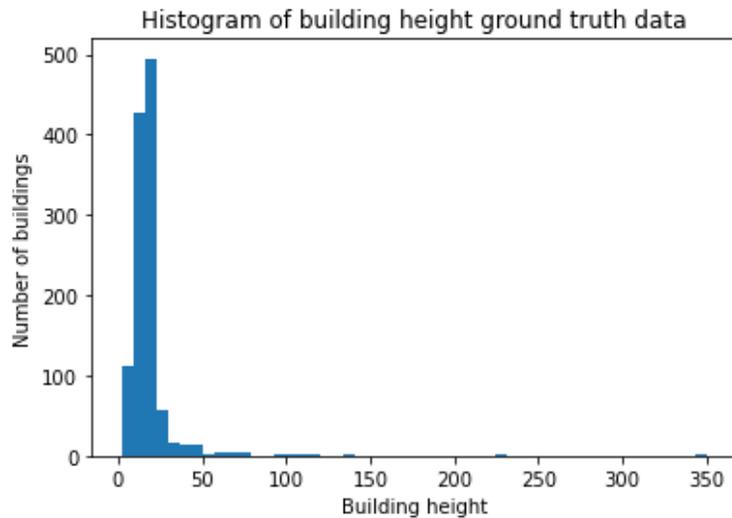
Figure 4: Histogram chart of building height ground truth data

Table 1: Number of buildings in the respective scenes

Scene ID	Number of included buildings
1	3785
2	2335
3	1869
4	1431
5	3432
6	2392
7	2530
8	509
9	355
10	1406
11	433
12	3723
13	3849
14	159
15	412
16	1324
Total	**29944**

The data set is provided as shapefile with 1411 records corresponding the building polygons, and the height is an attribute of each polygon. First, the shapefile data will be manually georeferenced to match the current image and angle from high-resolution satellite imagery from Google Earth. The data set will then be checked and filtered for polygons that contain no information or incorrect file format. After the ground truth data was preprocessed, a statistical evaluation of the height of the buildings was conducted in the ground truth dataset. The histogram chart of the ground-truth data is shown in Figure 4.

After statistically evaluating ground truth data, skyscrapers or buildings that are too low (<3m) will be considered as outlier and removed. After performing this step, the data set remains 1136 polygons. Of which, 70% of the data will be used as a training set (792 polygons) and 30% of the data will be used as a test set (344 polygons).

2.3 Building Footprint Extraction

The building footprint data on the study area is extracted manually, although manual extraction takes a lot of time, it guarantees higher accuracy than the use of classification and segmentation algorithms. The number of buildings in each scene is shown in Table 1. A total of 29,994 buildings are marked in the Cau Giay area. Of which, scene 13 has the most buildings with 3,849 building blocks and scene 14 has the least number of buildings with 159 building blocks. Extracted house foundation data on all 16 scenes is shown in Figure 5. Each scene is the same size (width and height of the image), but due to the different design and construction planning at each location, the house density in each scene are different.

2.4 Development of Building Height Regression Model

The height estimation method is shown in the Figure 5. Our proposed approach will have five main steps to estimated the building height from aerial images including (1) masking the input images, (2) key points extraction, (3) Bag of Word (BoW) feature creation, training the regression model, and (5) evaluation and validation.
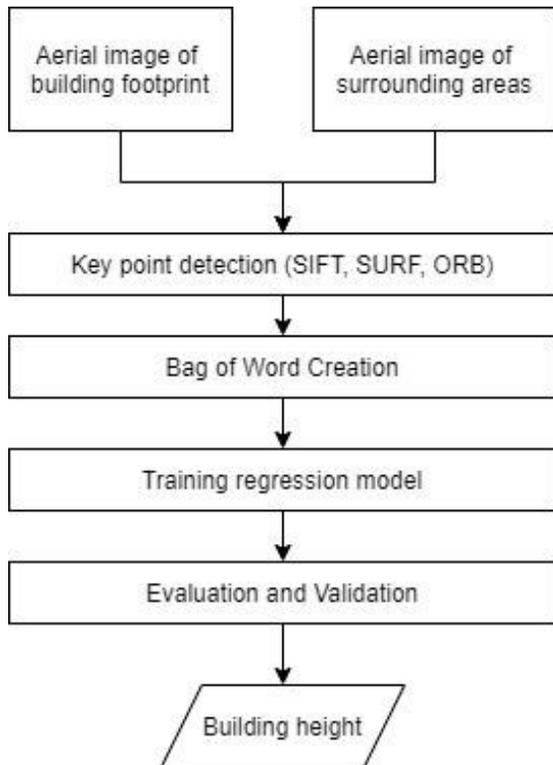
Figure 5: The flowchart of the proposed methodology for building height estimation

(1) First, the house foundation image will be extracted based on the manually collected shapefile. Here we distinguish the building footprint and the surrounding area into two separated images as in Figure 6a (building footprint) and Figure 6b (a buffer surrounding the building). As the surrounding area somehow correlated to the building height. For example, the height of the building which is closed to the main road versus a house which is located in the crowded residential area is much different. Base on that hypothesis, the authors have visually examined the extent to which buffer is the best fit for the study area ranging from 10m, 25m, 50m, and 100m, and the 25m seems to be the best choice for our region of interest.

(2) Then, respectively key point detectors including Scale Invariant Feature Transform (SIFT) (Lowe, 2004), Speed up Robust Feature (SURF) (Bay et al., 2008) and Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011) were applied to extract key points for each house and surrounding area. SIFT, SURF and ORB are reputed for each performance in image matching. SIFT is proposed by Lowe and has proven to be efficient in object recognition applications. However, SIFT still has some drawbacks related to computation

performance which affects the real-time application (Karami et al., 2017). SURF is also a point descriptor like SIFT but faster because of its quality reduction of detected points (Bay et al., 2008). ORB, which is proposed by Rublee and is an alternative of SURF and SIFT as these two descriptors are patented. SIFT, SURF and ORB has been employed in remote sensing application mostly for sea ice applications (Rothrock et al., 1999 and Liu et al., 2013). The reason that the key point detectors were used here is based on our assumption that roof shape and the surrounding buffer may partially represent for the average building height of a small appropriate area.

(3) After extracting the key points, BoW feature will be generated from the extracted key points in each image based on K-means as this model has gained a reputation in clustering problem, especially in BoW feature extraction. Here we applied the cross-validation strategy to optimize the hyper parameters for K-means cluster and also to avoid overfitting from the clustering process.

(4)-(5) Finally, the Random Forest model will be employed to train the height regression model from the BoW feature generated from the extracted key points. The cross-validation strategy with 3 folds is also applied for the training process for hyper parameters optimization and overfitting prevention. The height regression model will be evaluated through the mean absolute error (MAE) and root-mean-square error (RMSE) indicators.

2.5 3D Modelling
After the height data is estimated, the 3D city model for Cau Giay will be automatically constructed based on the workspace developed using FME software. FME is a powerful ETL (Extract, Transform and Load) software for converting and processing different data types and popular in GIS. In this study, we built the LoD1 3D model automatically with the input shapefile and the output data format is 3D-Tile b3dm. The buildings will be colored correspondingly for height classification in a 3D city model.

3. Results and Discussion
3.1 Building foot print extraction
House foundation data was extracted manually using ENVI 5.2 application. The result of 29944 foundations for 16 scenes is shown in Figure 7. It can be seen that the density distribution of the houses in the 16 scene is different according to the construction planning issued by the city.

a) Aerial image of building footprint
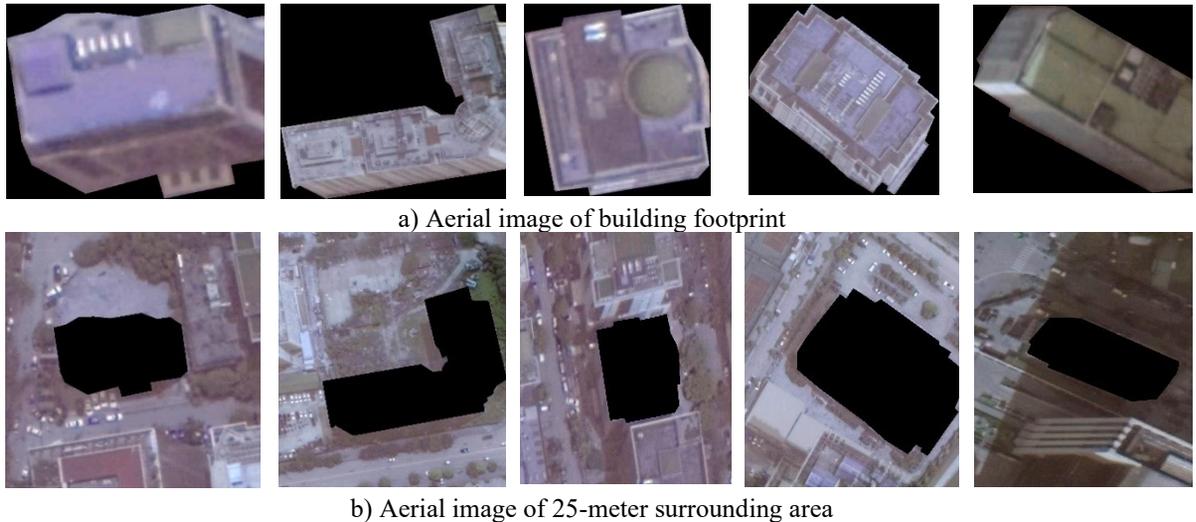

b) Aerial image of 25-meter surrounding area

Figure 6: Examples of building footprint (a) and surrounding area (b) are used to characterize a height estimation model

Table 2: The performance of height regression model in different cases

Key point detector	Aerial image of 25-meter surrounding area	Aerial image of building footprint	MAE (m)	RMSE
Case 1	SIFT	-	3.8	39.6
Case 2	SURF	-	3.7	35
Case 3	ORB	-	4.3	49.1
Case 4	-	SIFT	4.1	38.9
Case 5	-	SURF	4.3	41.2
Case 6	-	ORB	4.5	44.3
Case 7	SIFT	SIFT	3.7	37.3
Case 8	SIFT	SURF	3.6	34.6
Case 9	SIFT	ORB	3.8	37
Case 10	**SURF**	**SURF**	**3.5**	**33.6**
Case 11	SURF	ORB	3.7	34.6
Case 12	SURF	SIFT	3.7	34.9
Case 13	ORB	ORB	4.1	41.5
Case 14	ORB	SURF	4.1	39.7
Case 15	ORB	SIFT	4.1	41.5

3.2 Building height estimation

The results of the height estimates are described in Table 2. Where the individual use cases and the combination of the building footprint image and the surrounding area image, were evaluated with key point detectors. The K-means clustering model and the Random Forest regressor are optimized for the super-parameters through cross-validation method, the best model was obtained with number of clusters = 250, batch size = 200 for K-means. For the Random Forest model, the model gives the best results for the set of hyper parameters: max depth = 200 and random state = 50.

The results of the height estimation evaluation model through the two metrics MAE and RMSE are shown in Table 2. It can be seen that ORB gives the lowest result in all cases. With single image, extracted features from the 25-meter surrounding area gave better results compared to the ones extracted from building footprint image. SURF gives the best results when using the surrounding area image with MAE = 3.7 and RMSE = 35, while ORB gives the highest error result with MAE = 4.3 and RMSE = 49.1. For building footprint images, SIFT gave the lowest error results with MAE = 4.1 and RMSE 38.9, ORB still gave the highest error with MAE = 4.5 and RMSE = 44.3. For single image, SURF with surrounding image result in higher accuracy and lower error than SIFT with single image.
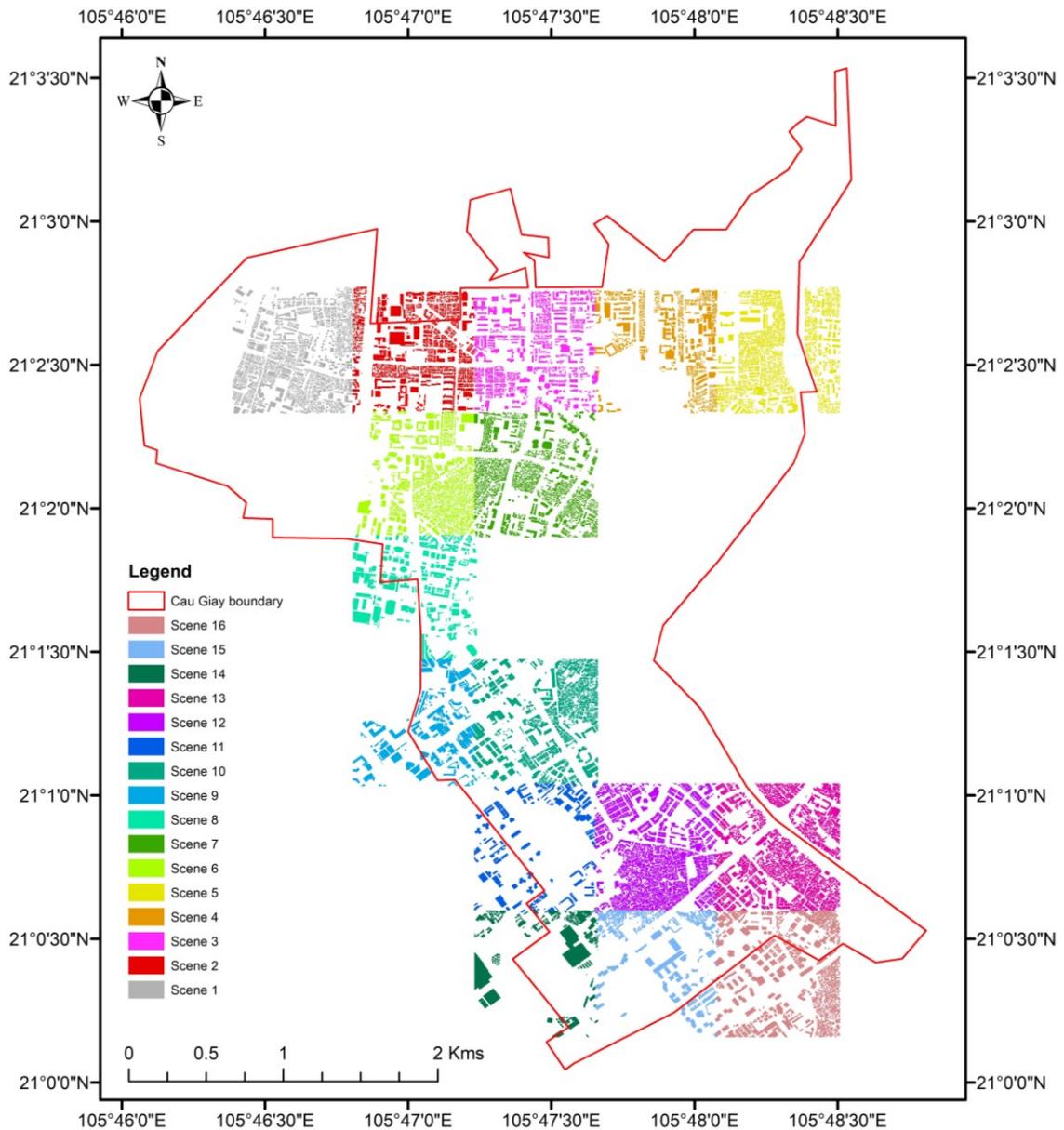
Figure 7: Building footprint data of 16 scenes covering Cau Giay district, Hanoi

For combining two types of image data, a total of 9 cases are designed to evaluate the effectiveness of using key point descriptor. Each pair of key point detectors in turn will be stitched together for the two images. The specific results are shown from Case 7 to Case 15 in Table 2. In general, combining feature sets extracted from two images minimizes model error compared to using a single image. It can be seen that, applying SURF to two types of images still gives the best results when achieving MAE =

3.5 and RMSE = 33.6. Meanwhile, the combination of ORB in the surrounding area image with other detectors in the building footprint all gave low results with MAE = 4.1 and RMSE ranging from 39.7 to 41.5. SIFT also outperforms SURF (MAE between 3.6 and 3.8 and RMSE between 34.6 and 37.3), but is much better than ORB when used for surrounding area images. Figure 8 is colored 2D building footprints that correspond to the building height range.
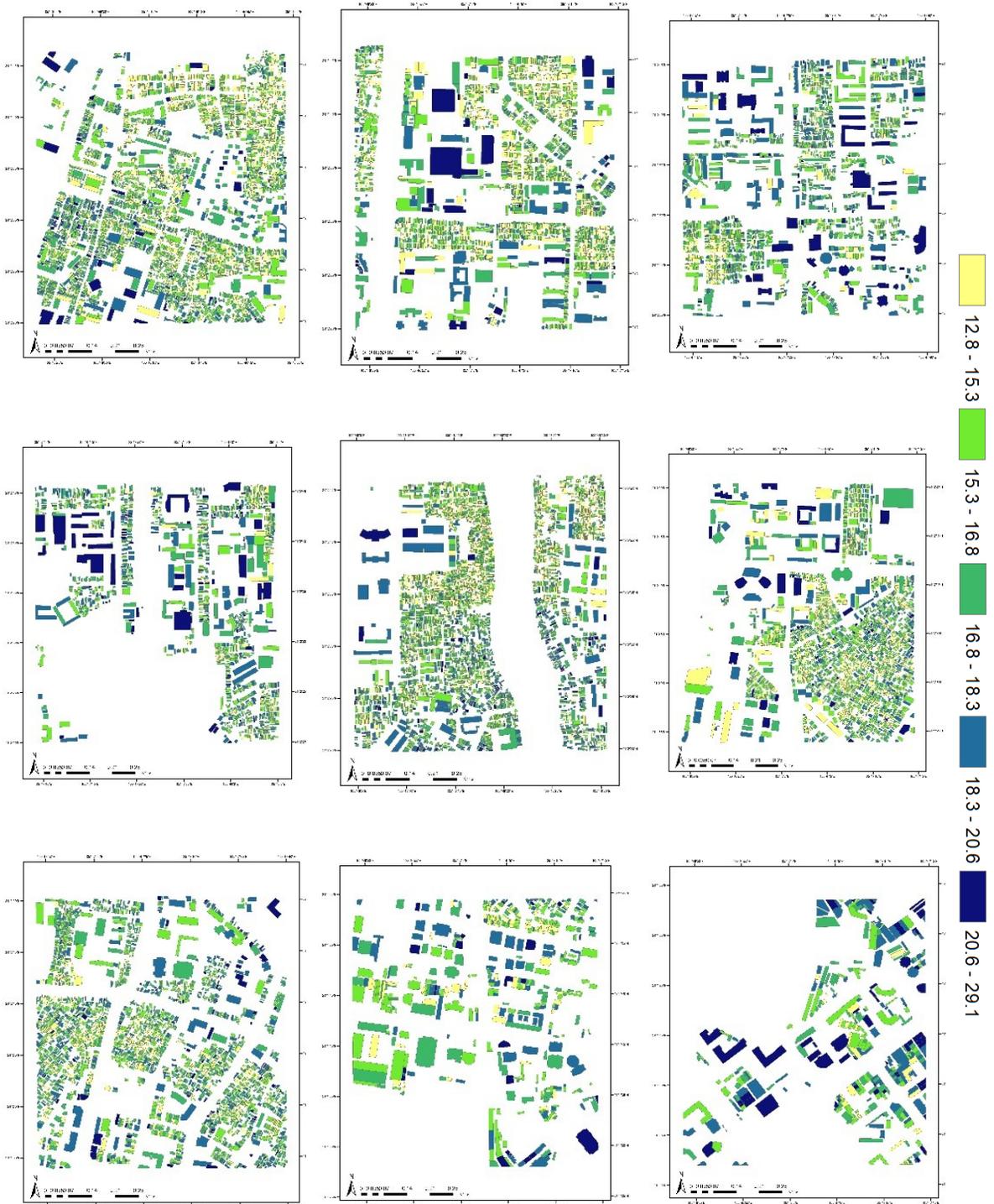
Figure 8. Building height estimation result for 16 scenes covering Cau Giay district
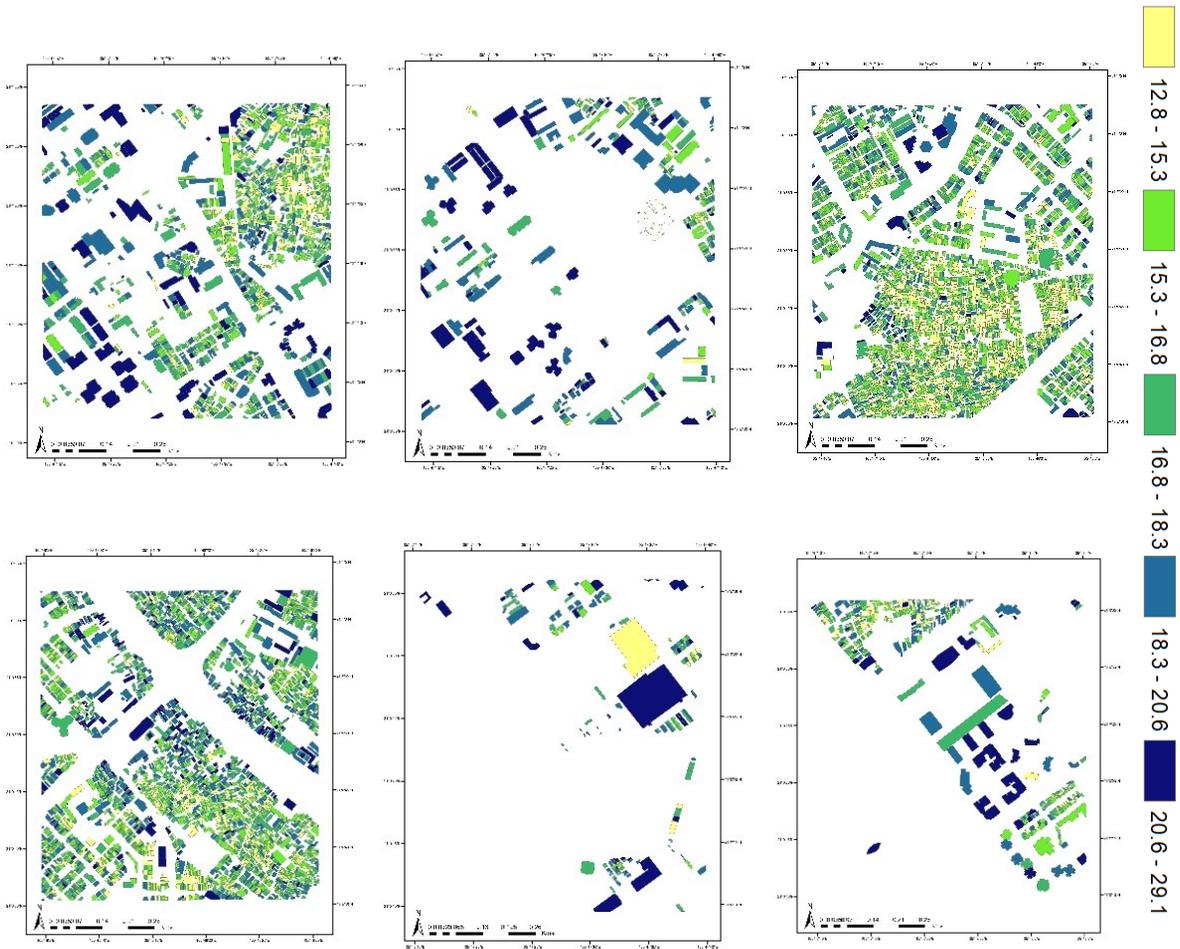(continue next page)

Figure 8: Building height estimation result for 16 scenes covering Cau Giay district
(continue from previous page)
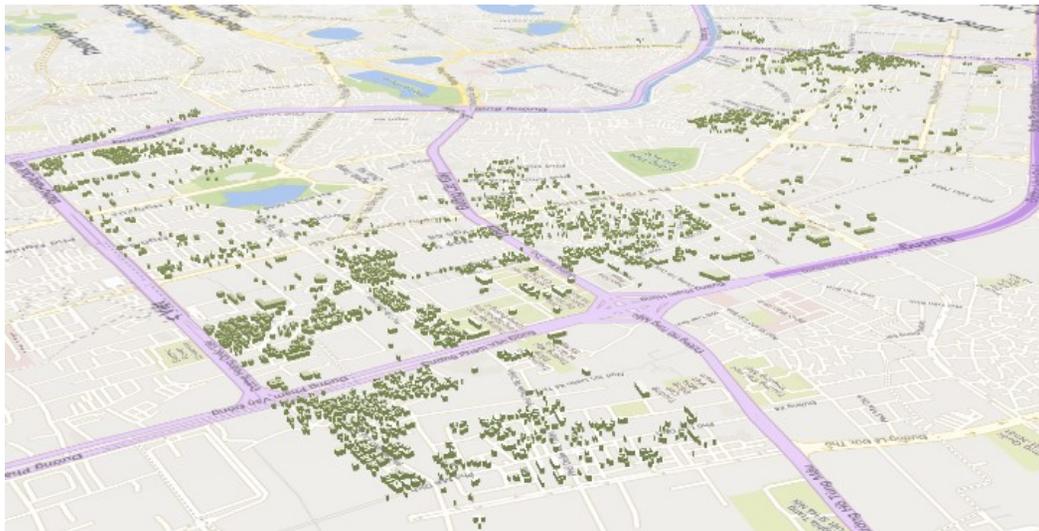
3.3 3D city modelling

After the height data is estimated, a 3D model for the entire city is constructed. The 3D city image is depicted as shown in Figure 9. Figure 9a shows the houses in the 12.8 - 15.3m height range, houses with a height between 15.3 and 16.8 are shown in Figure 9b. Similar to Figures 9c, 9d, 9e, 9f are 3D models with building height of 16.8 - 18.3m, 18.3 - 20.6m, 20.6m - 29.1m, and> 100m, respectively. Figure 9g shows all the houses with colors corresponding to the height range.

The number of buildings corresponds to the height class described in Table 3. It can be seen that most of the houses in the study area have a height between 12.8m and 18.3m, these are built with 3 to 6 floors. Houses taller than 18.3 - 20.6m account for about 18% of the total. Skyscrapers with a height > 100m make up the least, with total 44 houses across the district. It can be said that the number of skyscrape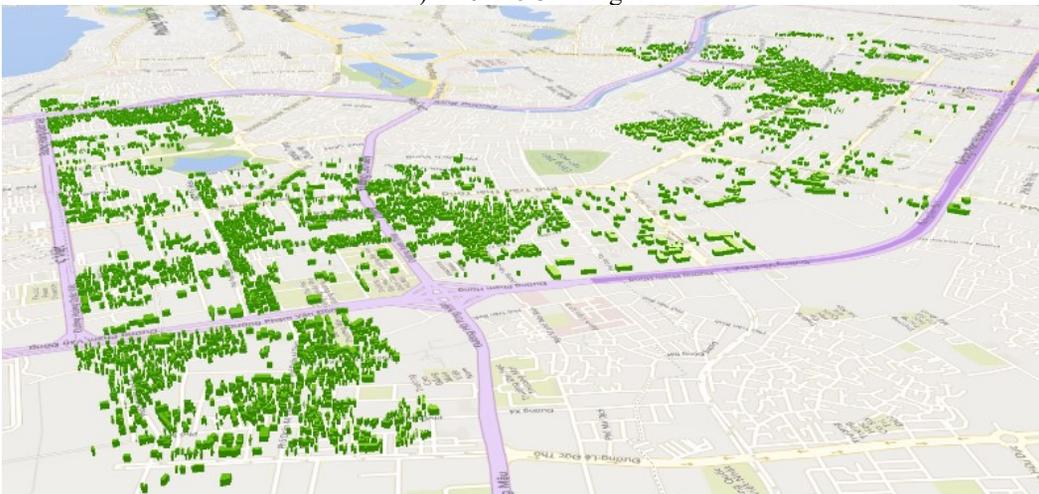rs and the height of buildings also represent the degree of urban development of an area. In this study, height data for skyscrapers was collected from open sources on the internet.
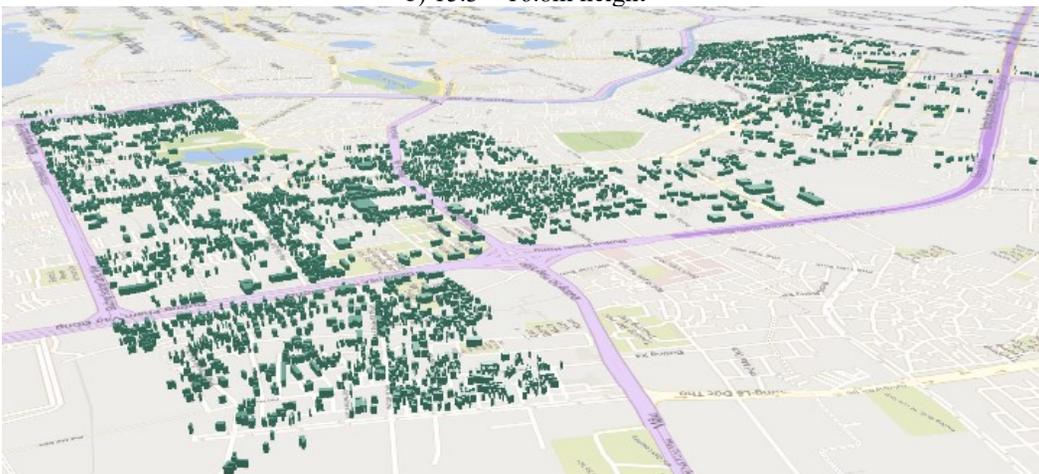
4. Conclusions

In this study, we proposed a method to build 3D city models from satellite images. First the house foundation data was extracted manually based on high resolution satellite images provided by Google Earth. Then, the surrounding 25-metter area image is extracted and used to extract the feature for the height regression model. SURF descriptor was used to extract key point points from two images and performed the best result compared with ORB and SIFT with MAE = 3.5 and RMSE = 33.6. This error result is accepted in the CityGML data standard for the 3D LoD1 building model. The machine learning model used in the height regression is Random Forest.
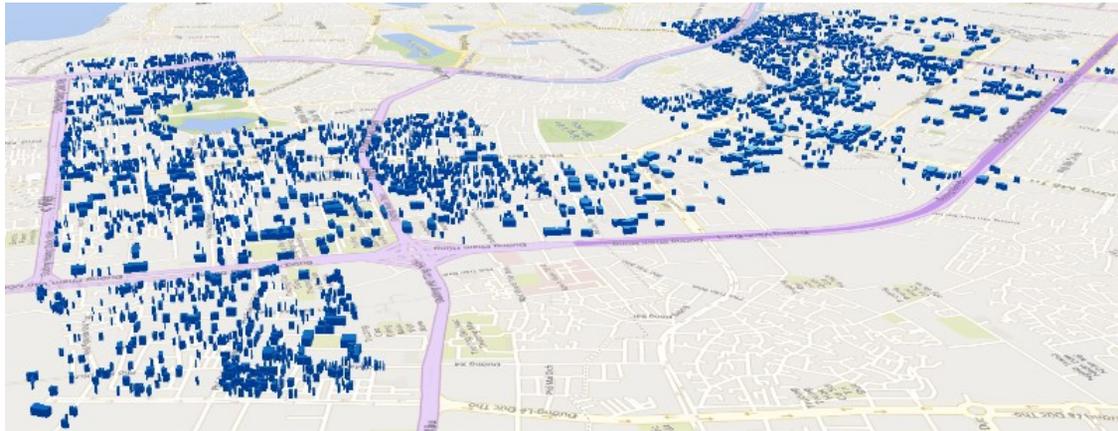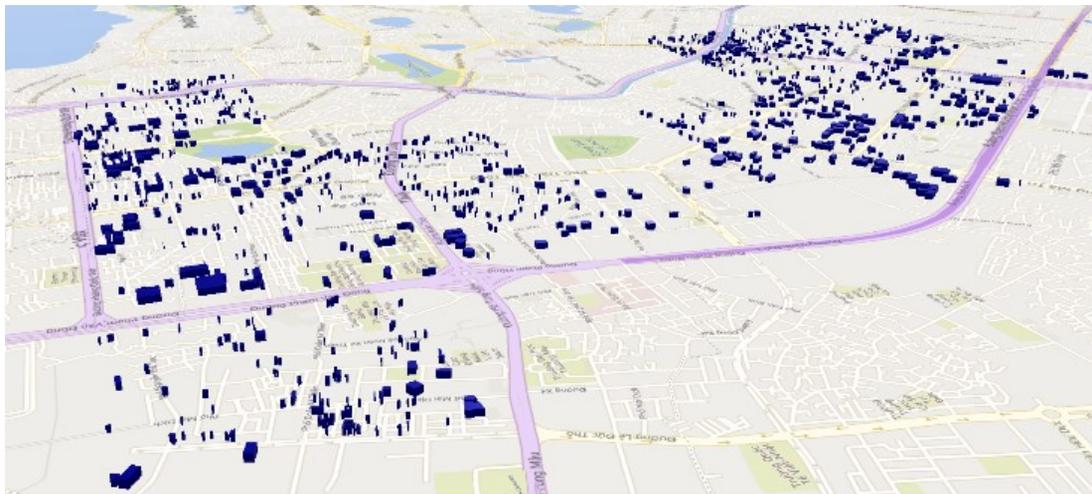
a) 12.8 - 15.3m height

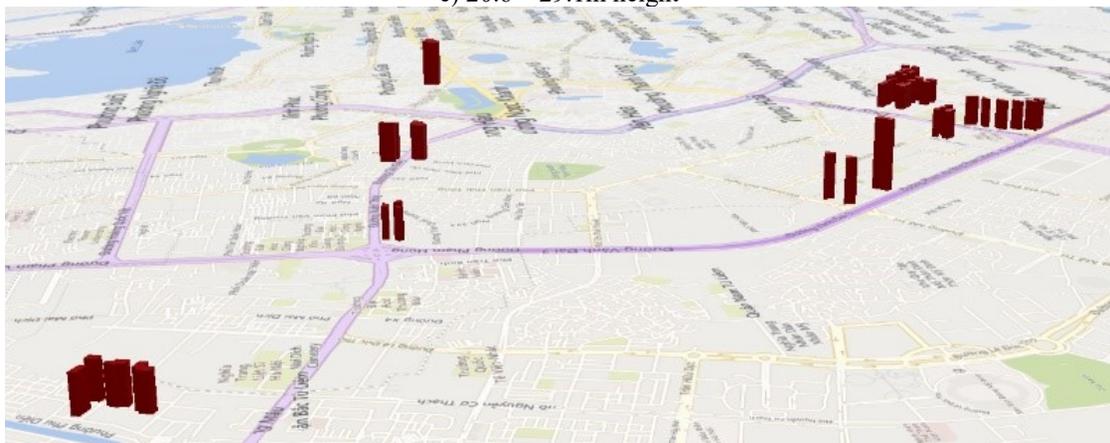b) 15.3 – 16.8m height

c) 16.8 – 18.3m height

Figure 9: 3D buildings in Cau Giay district, colors show the corresponding height classes (continne next page)

d) 18.3 – 20.6m height

e) 20.6 – 29.1m height

f) > 100m height

Figure 9: 3D buildings in Cau Giay district, colors show the corresponding height classes (continne next page)

g) All buildings are displayed in 3D with colors corresponding to the height range

Figure 9: 3D buildings in Cau Giay district, colors show the corresponding height classes
(continue from previous page)

Finally, the 3D-tile scheme is applied to build a 3D model for the study area – Cau Giay, the building will be colored corresponding to the six height levels > 100m, 20.6 - 29.1m, 18.3 - 20.6m, 16.8 - 18.3m, 15.3 - 16.8m, 12.8 - 15.3m.The proposed method helps to solve the problem of lack of input

data (building footprint and building height) to build a 3D city at the level of detail LoD1 when these data are not normally available in the developing city or country. The study still has a number of limitations including: the model building is not 100% automated while the building footprint data is still collected manually. The training data provided for the machine learning regression model is small (1136/29944 buildings ~ 3.7% of the data used for training), and the training data has not been stratified by height ranges in the study area. Therefore, estimating the error for actual building height estimation on the site may be even larger than MAE = 3.5 (m) and RMSE = 33.6.

In the future, we will continue to develop and improve the 3D modeling process using only high-resolution satellite image data. Specifically, the method of extracting house foundation data will continue to be developed from the old research of the group (Nguyen et al., 2019). In addition, the height data sampling strategy will also be appropriately designed and additional ground truth data will be collected to improve the efficiency of height estimation on this study area.

Acknowledgments

References

Anh, P., Chuc, M. D., Hung, B. Q. and Thanh, N. T. N., 2017, Development of Virtual Campus Using GIS Data and 3D GIS Technology: A Case Study for Vietnam National University, Hanoi. *Proceedings - 2017 9th International Conference on Knowledge and Systems Engineering,* DOI:10.1109/KSE.2017.8119472.

Anh, P., Thanh, N. T. N., Vu, C. T., Ha, N. V. and Hung, B. Q., 2019, Preliminary Result of 3D City Modelling for Hanoi, Vietnam. *NICS 2018 - Proceedings of 2018 5th NAFOSTED Conference on Information and Computer Science.* 294-299, DOI:10.1109/NICS.2018.8606867.

Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V., 2008, Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, Vol. 110(3). doi: 10.1016/j.cviu.2007.09.014.

Biljecki, F., Ledoux, H. and Stoter, J., 2017, Computers, Environment and Urban Systems Generating 3D City Models without Elevation Data. *Computers, Environment and Urban Systems*, Vol. 64, 1–18. doi: 10.1016/j.compenv-urbsys.2017.01.001.

Davydova, K., Cui, S. and Reinartz, P., 2016, Building Footprint Extraction from Digital Surface Models Using Neural Networks. *Proc. SPIE 10004, Image and Signal Processing for Remote Sensing XXII*. https://doi.org/10.1117/12.2240727.

Haithcoat, T. L., Song, W. and Hipple, J. D., 2001, Building Footprint Extraction and 3-D Reconstruction from LIDAR Data. *IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (Cat. No.01EX482)*. 958-964.

Irvin, R. B. and McKeown, D. M., 1989, Methods for Exploiting the Relationship Between Buildings and their Shadows in Aerial Imagery. *IEEE Transactions on Systems, Man and Cybernetics*. Vol. 19(6), doi: 10.1109/21.44071.

Kadhim, N. and Mourshed, M., 2018, A Shadow-Overlapping Algorithm for Estimating Building Heights from VHR Satellite Images. *IEEE Geoscience and Remote Sensing Letters*, Vol. 15(1). doi: 10.1109/LGRS.2017.2762424.

Karami, E., Shehata, M. and Smith, A., 2017, Image Identification Using SIFT Algorithm: Performance Analysis against Different Image Deformations. *ArXiv*. 1-4, https://arxiv.org/-ftp/arxiv/papers/1710/1710.02728.pdf.

Li, W., He, C., Fang, J., Zheng, J., Fu, H. and Yu, J., 2019, Semantic Segmentation-Based Building Footprint Extraction Using Very High-Resolution Satellite Images and Multi-Source GIS Data. *Remote Sensing*, Vol. 11(4). doi: 10.3390/rs11040403.

Liasis, G. and Stavrou, S., 2016a, Building Extraction in Satellite Images Using Active Contours and Colour Features. *International Journal of Remote Sensing*, Vol. 37(5). doi: 10.1080/01431161.2016.1148283.

Liasis, G. and Stavrou, S., 2016b, Satellite Images Analysis for Shadow Detection and Building Height Estimation. *ISPRS Journal of Photogrammetry and Remote Sensing,* Vol. 119, doi: 10.1016/j.isprsjprs.2016.07.006.

Mou, J. and Zhu, X. X., 2018, IM2HEIGHT: Height Estimation from Single Monocular Imagery via Fully Residual Convolutional-Deconvolutional Network. *Computer Vision and Pattern Recognition*. 1-13.

Liu, Z., Zhao, Z., Fan, Y. and Tian, D., 2013, Automated Change Detection of Multi-Level Icebergs near Mertz Glacier Region Using Feature Vector Matching. *Proceedings of the 2013 International Conference on Image Processing, Computer Vision, and Pattern Recognition.* 1-6, http://worldcomp-proceedings.com/proc/p2013/IPC3423.pdf.

Lowe, D. G. 2004, Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60(2), doi: 10.1023/B:VISI.0000029664.99615.94.

Nguyen, A., Luu, H., Phan, A., Bui, B. and Nguyen, T., 2019, Cau Giay: A Dataset for Very Dense Building Extraction from Google Earth Imagery. *Proceedings-2019 6th NAFOSTED Conference on Information and Computer Science, NICS.* DOI: 10.1109/NICS48868.2019.9023848.

Qi, F., Zhai, J. Z. and Dang, G., 2016, Building Height Estimation Using Google Earth. *Energy and Buildings*, Vol. 118. doi: 10.1016/j.enbuild.2016.02.044.

Rothrock, D. A., Yu, Y. and Maykut, G. A., 1999, Thinning of the Arctic Sea-Ice Cover. *Geophysical Research Letters*, Vol. 26(23). doi: 10.1029/1999GL010863.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G., 2011, ORB: An Efficient Alternative to SIFT or SURF. *Proceedings of the IEEE International Conference on Computer Vision*, 1-8, DOI: 10.1109/ICCV.2011.6126544.

Wang, O., Lodha, S. K. and Helmbold, D. P., 2006, *A Bayesian Approach to Building Footprint Extraction from Aerial LIDAR Data*. 1-8, DOI:10.1109/3DPVT.2006.9.

Zhao, K., Kang, J., Jung, J. and Sohn, G., 2018, Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 247-251. DOI:10.1109/CVPRW.2018.00045.

Zhao, Y., Qi, J. and Zhang, R., 2019, CBHE: Corner-Based Building Height Estimation for Complex Street Scene Images. *The Web Conference 2019 - Proceedings of the World Wide Web Conference,* DOI:10.1145/3-308558.3313394.