BUILDING CHANGE DETECTION FROM BITEMPORAL AERIAL IMAGES USING DEEP LEARNING

Shu Su [1, *], Takahiko Nawata [1], Takeru Fuse [1]

[1] Aero Asahi Corporation, Kawagoe, Saitama, 350-1165, Japan - (shuu-so, takahiko-nawata, takeru-fuse) @aeroasahi.co.jp

KEY WORDS: Building change detection, Aerial images, Digital surface models (DSMs), Building map, Deep learning, Urban area

ABSTRACT:

Automatic building change detection has become a topical issue owing to its wide range of applications, such as updating building maps. However, accurate building change detection remains challenging, particularly in urban areas. Thus far, there has been limited research on the use of the outdated building map (the building map before the update, referred to herein as the old-map) to increase the accuracy of building change detection. This paper presents a novel deep-learning-based method for building change detection using bitemporal aerial images containing RGB bands, bitemporal digital surface models (DSMs), and an old-map. The aerial images have two types of spatial resolutions, 12.5 cm or 16 cm, and the cell size of the DSMs is 50 cm × 50 cm. The bitemporal aerial images, the height variations calculated using the differences between the bitemporal DSMs, and the old-map were fed into a network architecture to build an automatic building change detection model. The performance of the model was quantitatively and qualitatively evaluated for an urban area that covered approximately 10 km^2 and contained over 21,000 buildings. The results indicate that it can detect the building changes with optimum accuracy as compared to other methods that use inputs such as i) bitemporal aerial images only, ii) bitemporal aerial images and bitemporal DSMs, and iii) bitemporal aerial images and an old-map. The proposed method achieved recall rates of 89.3%, 88.8%, and 99.5% for new, demolished, and other buildings, respectively. The results also demonstrate that the old-map is an effective data source for increasing building change detection accuracy.

1. INTRODUCTION

Over the last decade, with advances in computer vision techniques, building change detection has emerged as a promising field of research in the areas of photogrammetry and remote sensing. This general increase in interest may be attributed to its wide range of applications. One such application is updating building maps. In a building map, the building boundary is generally delineated based on orthorectified aerial images. For buildings whose boundaries are difficult to judge in this manner (such as a building covered by trees), field surveys are conducted. Traditionally, building change detection was performed manually by comparing aerial images from different time periods. Owing to the tedious and time-consuming nature of this task, researchers have developed automatic detection techniques. However, accurate building change detection remains challenging because of differences in the images used arising from differences in the cameras, atmospheric conditions, and solar angles. Previous studies on building change detection have primarily used bitemporal aerial images, focusing on their spectral information and color variations (Bourdis et al., 2011). However, the methods involved can produce several errors, including misdetection (i.e., a changed building going undetected) and overdetection (i.e., an unchanged building being detected as changed, or several changes attributed to a non-building area). The development of digital surface models (DSMs) has proven to be effective in improving the accuracy of building extraction and change detection (Murakami et al., 1999; Matikainen et al., 2003; Rottensteiner et al., 2007; Matikainen et al., 2010; Tian et al., 2014). Height variations represent robust feature information, which contributes to building change detection. However, it is challenging to distinguish between building changes and non-building changes

such as those in trees and vegetation. Furthermore, it is difficult to determine the height of a building if the building is partially obstructed by trees or the shadows of nearby buildings.

Recently, deep learning techniques have garnered considerable attention for achieving satisfying results in several classification problems. Researchers have made efforts to adopt deep learning techniques to solve the problems of automatic building change detection. Convolutional neural networks, which have proven to be effective in identifying objects based on their appearance variations, have received particular interest in building change detection (Daudt et al., 2018; Lim et al., 2018; Maltezos et al., 2018; Pang et al., 2018; Ji et al., 2019). However, further research on automatic building change detection is required to address the challenges that exist in urban areas, such as multiple small buildings being in close proximity, buildings being partially or wholly obstructed from view, and buildings that are complex in shape. The data sources utilized for building change detection techniques in previous works are categorized into the following three types: i) airborne or satellite imagery data, ii) three-dimensional data (i.e., DSMs and digital terrain models (DTMs)), and iii) both i) and ii) (Maltezos et al., 2018). However, several researchers have considered the possibility of using an outdated building map (the building map before an update), referred to herein as the old-map, as an input data source to improve the accuracy of building change detection. In Japan, most municipal governments acquire aerial images and update building maps every year or three years. Therefore, the aerial images and building map for the previous period can be obtained at a relatively low cost.

This work investigates the potential of an old-map for improving the accuracy of building change detection. A new method is presented for building change detection that uses bitemporal aerial images, bitemporal DSMs, and an old-map, along with deep learning techniques. The performance was evaluated for an urban area that covered approximately 10 km^2

---

* Corresponding author

and contained over 21,000 buildings. The proposed method was compared to methods that rely on the following inputs: i) bitemporal aerial images only, ii) bitemporal aerial images and bitemporal DSMs, and iii) bitemporal aerial images and an old-map.

2. DATA DESCRIPTION

In this study, two urban areas located in different provinces of Japan were selected to demonstrate the proposed method. One area, the training area, was used to generate training data to build the automatic building change detection model, whereas the other, the test area, was used to evaluate the method. The training area was over 50 km^2 in size and contained approximately 110,000 buildings. The test area was approximately 10 km^2 in size and contained over 21,000 buildings. Both the training area and test area had typical urban features, such as small buildings in close proximity, buildings partially or wholly shaded by nearby buildings, buildings with complex shapes, and numerous plants and roads.

2.1 Input data sources

An overview of the input data sources is given in Table 1.

Area	Data	Year	Spatial Resolution
Training	Old-ortho	2013	16 cm
	New-ortho	2015	12.5 cm
	Old-DSM	2013	50 cm
	New-DSM	2015	50 cm
	Old-map	2013	-
Test	Old-ortho	2018	12.5 cm
	New-ortho	2019	12.5 cm
	Old-DSM	2018	50 cm
	New-DSM	2019	50 cm
	Old-map	2018	-

Table 1. Overview of input data sources

2.1.1 Bitemporal aerial images: Bitemporal aerial images of the training area were acquired in 2013 and 2015. Aerial images of the test area were acquired in 2018 and 2019. Ortho images were derived from the aerial images. The ortho images generated from the earlier and later aerial images are referred to herein as old-ortho and new-ortho images, respectively; all of them are not true ortho images, that is, not all vertical features were re-projected into ortho images. The spatial resolution of the old-ortho images of the training area (16 cm/pixel) differs from that of other ortho images (12.5 cm/pixel).

2.1.2 Bitemporal DSMs: DSMs were obtained from the aerial images using a stereo matching technique; those obtained from the earlier and later aerial images are referred to herein as old-DSM and new-DSM, respectively. DSMs obtained in text format consisted of a list of rows with X, Y, and Z coordinates. The spatial resolution of the DSMs was 50 cm × 50 cm for each cell.

2.1.3 Old-map: The old-map used in this study was obtained for the same period as the earlier aerial images. Each polygon in the old-map represents a boundary of a building.

2.2 Reference map

By comparing old-ortho and new-ortho images, a building change map (referred to herein as a reference map) was manually created based on the old-map. The objects in the reference map were categorized into three groups: new buildings, demolished buildings, and other buildings. Other buildings included unchanged buildings, new building parts, and demolished building parts. New buildings were added and labeled as "new buildings", demolished buildings were labeled as "demolished buildings", and other buildings were labeled as "other buildings". Reference maps were created for both the training and test areas. The reference map of the training area was used to establish the ground truth assigned to the training dataset, whereas that of the test area was utilized to evaluate the performance of the proposed method. The reference map of the training area consisted of 3,298 new buildings, 2,091 demolished buildings, and 108,593 other buildings. The reference map of the test area consisted of 193 new buildings (150 of them greater than 20 m^2 in size), 189 demolished buildings (143 of them greater than 20 m^2 in size), and 21,616 other buildings. The small buildings (less than 20 m^2 in size), which were mostly warehouses, were not considered in this study; only new buildings and demolished buildings greater than 20 m^2 in size were considered. Thus, the performance of the proposed method was evaluated using the 150 new buildings, 143 demolished buildings, and 21,616 other buildings in the test area.

3. BUILDING CHANGE DETECTION METHOD

The following are the processing steps. First, the bitemporal ortho images were resampled to achieve the same spatial resolution. Second, color correction of the old-ortho and new-ortho images of both the training area and test area was conducted. Third, height variations were calculated from the differences between the old-DSM and new-DSM. Fourth, the old-map image was exported from the old-map. Finally, the old-ortho images, new-ortho images, height variations, and old-map images were fed into a network architecture to train the building change detection model that was applied to the test area. More details on these process steps are provided below.

3.1 Resampling

All of the ortho images were adjusted to the same spatial resolution as the old-ortho images of the training area (16 cm/pixel).

3.2 Ortho color correction

Significant differences in color variations in the old-ortho and new-ortho images were observed, which may have been caused by differences in the cameras, atmospheric conditions, and solar angles. To eliminate potential disturbances that these factors may cause to the performance of the building change detection method, the mean and standard deviation of each channel (RGB) were calculated for both the old-ortho and new-ortho images. The RGB value of each pixel in the old-ortho and new-ortho images was transformed using Equation 1.

$$y_i = \frac{x_i - m_i}{s_i} \qquad (1)$$

Here,
i = red, green, or blue channel
x_i = value of the channel
m_i = mean value of the channel
s_i = standard deviation of the channel
y_i = value of the channel after correction

3.3 Height variations

Height variations were obtained from value differences between the old-DSM and new-DSM, which were calculated by subtracting old-DSM values from new-DSM values and approximating each difference to the nearest integer. The height variation was read into the NumPy array at the same spatial resolution as the ortho image (16 cm/pixel). Compared to directly feeding the old-DSM and new-DSM into the network, height variations can reduce the amount of information given to the network.

3.4 Old-map image

The old-map image was exported and saved as a grayscale image. The areas within the building boundaries were white, and non-building areas were black.

3.5 Training/test phase

The training area was divided into blocks of 256×256 pixels without overlap. For each block, the old-ortho image, new-ortho image, height variation, and old-map image were considered as a training data sample set. Because of imbalances in the numbers of new, demolished, and other buildings, new and demolished building samples were augmented by cropping 256×256 pixels around the centers of new and demolished building polygons. The centers of these buildings were obtained from the reference map. A total of 37,252 training samples were generated, and a ground truth value was assigned to each training dataset created. Ground truth was transformed from the reference map, with red, blue, white, and black labels for the areas within a new building boundary, a demolished building boundary, any other building boundary, and a non-building area, respectively. The test area was divided into blocks of 256×256 pixels with a forward and side overlap of 56 pixels to obtain a seamless tiling of the predicted output.

U-Net is regarded as a remarkable, successful, and popular network architecture for semantic segmentation (Ronneberger et al., 2015). In this study, a U-Net-based encoder–decoder network architecture was designed, as illustrated in Figure 1. The encoder extracts multi-scale features, and the decoder uses skip-connections from the encoder to produce a more accurate localization of building boundaries. The encoder primarily consists of three blocks. Each block consists of two convolution layers (each activated by the tanh method) and a 2×2 max pooling operation with a stride of two for downsampling. The decoder also primarily consists of three blocks, each block having an upsampling of the feature map followed by a convolution layer that halved the number of feature channels, a concatenation with skip-connections from the encoder, and two

convolution layers (each activated using the tanh method). At the final layer, a 1×1 convolution (activated using the softmax method) is used to map each 64-component feature vector to the desired four classes (new, demolished, other, and non-building area). Batches of size 256×256 pixels cropped from old-ortho images, new-ortho images, height variations, and old-map images of the training area were fed into the network to build the model. Each batch consisted of eight channels, including an old-ortho image (three channels), a new-ortho image (three channels), height variation expressed as a NumPy array (one channel), and an old-map image (one channel). After the model was obtained, batches of size 256×256 pixels cropped from old-ortho images, new-ortho images, height variations, and old-map images of the test area were used as inputs to the model for predictive purposes. The size of the prediction results for each block of test data was 256×256 pixels. The predictions were merged in a large raster format with the test area as the region. During the merging process, the center region (200×200 pixels) within each prediction result was used for validation, while the marginal region was unused. To evaluate the performance of the proposed method, the raster format data were polygonized as polygon data in a shapefile format. The product of this step is referred to as a prediction map. Each polygon had classified information inherited from the colors in the prediction results. The colors red, blue, and white denoted a new building, a demolished building, and any other type of building, respectively. To eliminate the noise, the prediction map masked all polygons with area smaller than 3 m^2.

4. EXPERIMENTS AND RESULTS

Table 2 shows the three methods selected for comparison with the proposed building change detection method. The focus of the comparison was on evaluating the effectiveness of different types of input data sources. In the proposed method, old-ortho images, new-ortho images, height variations, and an old-map were used as input. In the O method, the input data consisted of only old-ortho and new-ortho images. In the O+H method, the input data consisted of old-ortho images, new-ortho images, and height variations. In the O+M method, the input data included old-ortho images, new-ortho images, and an old-map.

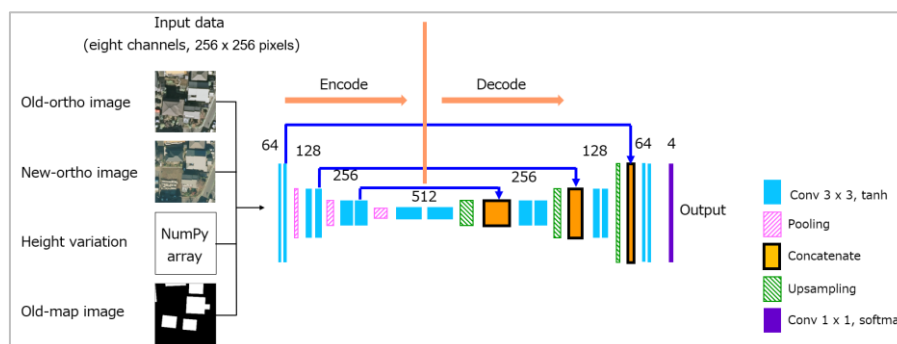| Method	Bitemporal ortho images	Height variations	Old-map
O	✔		
O+H	✔	✔	
O+M	✔		✔
Proposed	✔	✔	✔

Table 2. Input data types for comparison of methods

Figure 1. Network architecture

For all methods, the same training data and test data were used. The network architecture was the same in each case. The network was trained for 50 epochs using the Adadelta method (Zeiler, 2012) as the optimization algorithm for each method. All of the methods were trained on a workstation equipped with four NVIDIA GeForce RTX2080 Ti (11 GB) GPUs. Two methods were processed at a time; the training process for each method was allocated to two GPUs. The time required for each training process was approximately 13 hours.

4.1 Evaluation

The quantitative and qualitative evaluations of the methods were based on comparison with the reference map. The accuracy of the building change detection accomplished by each method was assessed using building-based accuracy measures. A confusion matrix was defined for new buildings, as shown in Table 3.

Actual / Predicted	New	Not New
New	TP	FP
Not New	FN	TN

Table 3. Confusion matrix for new buildings

True positive (TP) is the number of buildings if there was an overlap between new buildings in the prediction map and those in the reference map. False positive (FP) is the number of overdetections, that is, new buildings detected by the method that do not overlap with any new buildings in the reference map. False negative (FN) is the number of misdetections, that is, new buildings in the reference map that the method failed to accurately detect. The confusion matrices for demolished buildings and other buildings were defined in the same manner as that for new buildings.

The building detection accuracy achieved for each category of buildings (new, demolished, and other) was assessed using the precision, recall, and F2-score measures given by Equation 2.

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$
$$F2 - score = (1 + \beta^2) \frac{Precision \times Recall}{(\beta^2 \times Precision) + Recall}$$

(2)

Here,
TP = true positive
FP = false positive
FN = false negative
$\beta = 2$

A high recall value indicates a low misdetection rate, whereas a high precision indicates a low overdetection rate. The F2-score is a metric that combines recall and precision using the harmonic mean. The value of β reflects the ratio of the influences of precision and recall; for example, a β value of 2 indicates that the recall has more influence than the precision. This study was more focused on recall (which reflects the rate of misdetection) than precision (which reflects the rate of overdetection). This is because when the results of building change detection are used to update the old-map, overdetection can be corrected easily through subsequent human inspection, whereas misdetection requires the entire area to be checked to find all the buildings that have changed.

To evaluate the ability of each method to detect changes in new and demolished buildings, the average precision, recall, and F2-score were calculated using the micro-average method based on Equation 3. The values of the parameters referred to here as Aprecision, Arecall, and AF2-score, respectively. Aprecision and Arecall indicate the ability of a method to detect real changes and all real changes, respectively. The AF2-score is a metric that combines Aprecision and Arecall using the harmonic mean setting $\beta = 2$.

$$Aprecision = \frac{TP_{new} + TP_{demolished}}{TP_{new} + TP_{demolished} + FP_{new} + FP_{demolished}}$$
$$Arecall = \frac{TP_{new} + TP_{demolished}}{TP_{new} + TP_{demolished} + FN_{new} + FN_{demolished}}$$
$$AF2 - score = (1 + \beta^2) \frac{Aprecision \times Arecall}{(\beta^2 \times Aprecision) + Arecall}$$

(3)

Here,
TP_{new} = true positive for new building
$TP_{demolished}$ = true positive for demolished building
FP_{new} = false positive for new building
$FP_{demolished}$ = false positive for demolished building
FN_{new} = false negative for new building
$FN_{demolished}$ = false negative for demolished building
$\beta = 2$

The optimal value for precision, recall, F2-score, Aprecision, Arecall, and AF2-score is 1, i.e., when they are all equal to 1, no misdetection or overdetection occurs.

4.2 Results and discussion

The confusion matrix results for each building class for each method are shown in Table 4.

Method	Class	TP	FN	FP
O	New	126	24	263
	Demolished	87	56	184
	Other	18680	2936	2814
O+H	New	131	19	363
	Demolished	107	36	190
	Other	18606	3010	2469
O+M	New	127	23	192
	Demolished	108	35	104
	Other	21539	77	313
Proposed	New	134	16	206
	Demolished	127	16	182
	Other	21510	106	217

Table 4. Results for each method and building class

The calculated values of Aprecision, Arecall, AF2-score, precision, recall, and F2-score for each building class (new, demolished, and other) according to Table 4 are plotted in Figure 2.

4.2.1 Quantitative evaluation

As shown in Figure 2(a), the proposed method achieved a higher AF2-score (71.7%) than the other methods. This implies that it demonstrated optimum performance in detecting building changes, with an optimal balance of lower misdetection and lower overdetection rates. The O+M method achieved the second highest AF2-score of 69.0%, the O+H method achieved the third highest AF2-score of 60.6%, whereas the O method achieved the lowest AF2-score of 58.1%. Thus, it can be confirmed that neither DSMs nor old-map can account for the accuracy of the building change detection task. The proposed method also yielded the highest Arecall (Figure 2(a)), $recall_{new}$ (Figure 2(b)), and $recall_{demolished}$ (Figure 2(c)) scores of 89.1%, 89.3%, and 88.8%, respectively. This indicates that the proposed method achieved the lowest misdetection rate in the building change detection process.

To evaluate the impact of the old-map, the following pairwise comparisons of the methods without and with the old-map were conducted.
(1) O method vs. O+M method
The O+M method achieved better results than the O method by all measures (Figure 2(a)). The Aprecision, Arecall, and AF2-

score values for the O+M method were 44.3%, 80.2%, and 69.0%, respectively, whereas those for the O method were 32.2%, 72.7%, and 58.1%, respectively.
(2) O+H method vs. proposed method
The proposed method achieved better results than the O+H method by all measures (Figure 2(a)). The Aprecision, Arecall, and AF2-score values for the proposed method were 40.2%, 89.1%, and 71.7%, respectively, whereas those for the O+H method were 30.0%, 81.2%, and 60.6%.

As shown in Figure 2(d), the respective precision, recall, and F2-score values for the O+M method and the proposed method are significantly better than those for the O and O+H methods. This confirms the effectiveness of using the old-map as input for the building change detection task.

Additionally, the effect of using height variation information was compared with that of using the old-map. The O+M method achieved the highest Aprecision, as well as a higher AF2-score than the O+H method; the O+H method achieved a higher Arecall, but the difference was not significant (Figure 2(a)). These results suggest that the old-map is more useful than DSM data.
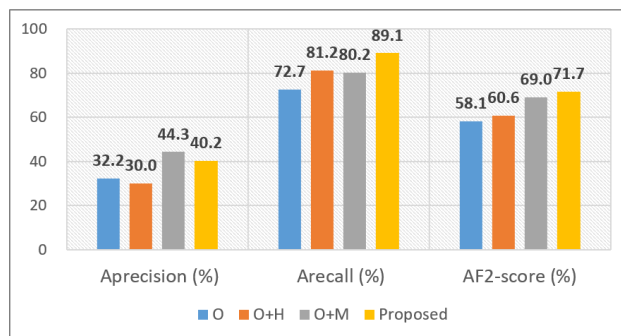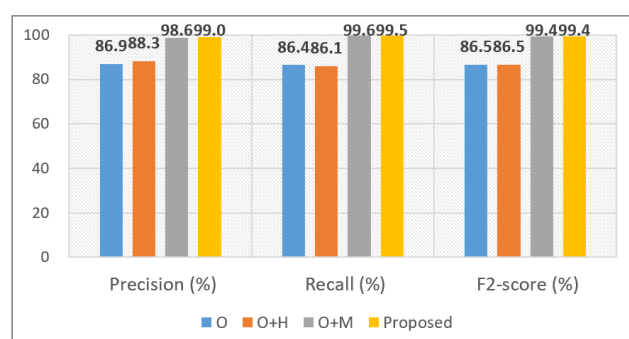

(a) Results for Aprecision, Arecall, and AF2-score
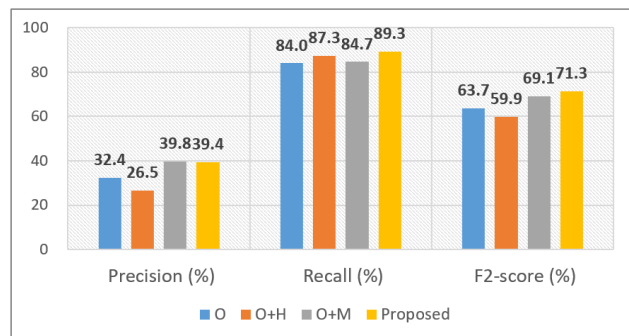
(b) Precision, recall, and F2-score for new building detection
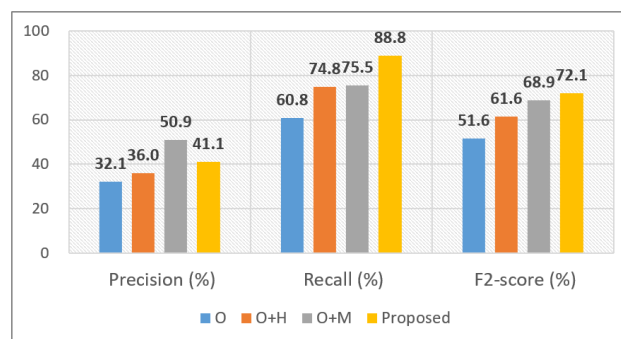
(c) Precision, recall, and F2-score for demolished building detection

(d) Precision, recall, and F2-score for other building detection

Figure 2. Quantitative results for the different methods

4.2.2 Qualitative evaluation

To qualitatively evaluate the results, sample automatic building change detection results obtained using the different methods were plotted, as shown in Figure 3. The size of the subarea shown in the figure is 100 m × 100 m. The red, blue, white, and black regions denote new buildings, demolished buildings, other buildings, and non-building areas, respectively. This plot confirms that the proposed method outperforms the other methods. The results obtained using the proposed method (Figure 3(g)) are similar to the reference map (Figure 3(c)). In particular, the boundary regions of demolished buildings and those of other buildings are almost identical. This is because the boundary information of demolished and other buildings was embedded in the old-map, which was used as an input. The O+M method (Figure 3(f)) effectively obtained the boundary regions of other buildings and yielded results almost identical to the reference map. It was also able to obtain the boundary regions of new buildings. However, its performance was comparatively poor for demolished buildings. As shown in Figure 3(f), the boundary of the demolished building in the upper left part of the image was extracted well (green frame), but a large portion of that of the demolished building at the center was missed (yellow frame). This may be because the O+M method does not use DSM data. As the color of that region in the new-ortho image is similar to the color of the roofs of some buildings, the O+M method misdetected the existence of those buildings. The prediction results for the O+H method (Figure 3(e)) include a non-existent new building (yellow frame) and a non-building marked as demolished (green frame). Compared to the proposed method, the O+H method performed worse in extracting building boundaries, particularly for other buildings. It is evident that the results of the O method (Figure 3(d)) are the least accurate of all in terms of new (yellow frame), demolished (green frame), and other buildings.

5. CONCLUSIONS

In this paper, a building change detection method that uses bitemporal aerial images, bitemporal DSMs, and an old-map is proposed. Batches of 256 × 256 pixels cropped from old-ortho images, new-ortho images, height variations, and old-map images were fed into a network based on the U-Net architecture to obtain a building change detection model. The performance of the proposed method was evaluated for an urban area of approximately 10 km², with 150 new buildings (greater than 20 m²), 143 demolished buildings (greater than 20 m²), and 21,616 other buildings. The proposed method achieved recall rates of 89.3%, 88.8%, and 99.5% for new, demolished, and other buildings, respectively. The results showed that the proposed method has low misdetection. The method over-detected 206 new, 182 demolished, and 217 other buildings (Table 4). However, these overdetections only amounted to a small percentage of the total of over 21,000 buildings. Furthermore, overdetections can be corrected easily by subsequent human inspection. Compared to the O, O+H, and O+M methods, the proposed method achieved the optimum, most balanced average quality rate. The results showed that the proposed method is suitable for building change detection tasks and demonstrated the effectiveness of the old-map as a data source for improving the accuracy of building change detection. In future work, we will optimize the hyper-parameters of the U-Net network, such as activation functions, and perform further comparisons with other methods that use inputs such as i) bitemporal DSMs and an old-map and ii) new-ortho images, bitemporal DSMs, and an old-map.
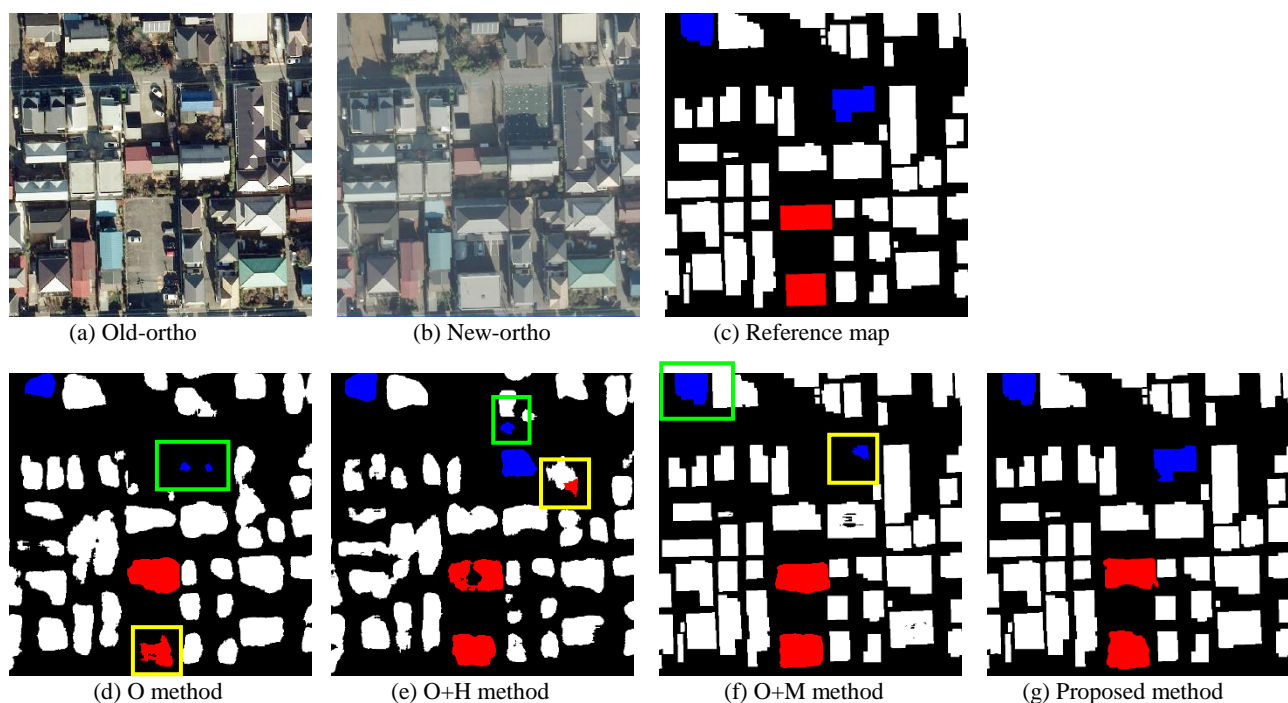
ACKNOWLEDGEMENTS

Figure 3. Examples of building change detection using different methods (100 m × 100 m)
(red: new; blue: demolished; white: other buildings; black: non-building)

REFERENCES

Bourdis, N., Marraud, D., Sahbi, H., 2011. Constrained optical flow for aerial image change detection. IGARSS, Vancouver, Canada.

Daudt, R., Saux, B., Boulch, A., 2018. Fully Convolutional Siamese Networks for Change Detection. ICIP 2018, Athens, Greece.

Ji, S.; Shen, Y.; Lu, M.; Zhang, Y., 2019. Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Samples. Remote Sens. 1343.

Lim, K., Jin, D., Kim, C., 2018. Change Detection in High Resolution Satellite Images Using an Ensemble of Convolutional Neural Networks. APSIPA ASC 2018, Honolulu, Hawaii.

Maltezos, E., Ioannidis, C., Doulamis, A. and Doulamis, N., 2018. Building Change Detection using Semantic Segmentation on Analogue Aerial Photos. FIG Congress 2018, Istanbul, Turkey.

Matikainen, L., Hyyppä, J., Hyyppä, H., 2003. Automatic detection of buildings from laser scanner data for map updating. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Dresden, Germany, Vol. XXXIV, Part 3/W13, pp. 218-224.

Matikainen, L.; Hyyppä, J.; Ahokas, E.; Markelin, L.; Kaartinen, H., 2010. Automatic Detection of Buildings and Changes in Buildings for Updating of Maps. Remote Sens. 1217-1248.

Murakami, H., Nakagawa, K., Hasegawa, H., Shibata, T., Iwanami, E., 1999. Change detection of buildings using an airborne laser scanner. ISPRS Journal of Photogrammetry and Remote Sensing, 54, pp. 148-152.

Pang, S., Hu, X., Cai, Z., Gong, J. and Zhang, M., 2018. Building Change Detection from Bi-Temporal Dense-Matching Point Clouds and Aerial Images. Sensors (Basel).

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351, 234-241.

Rottensteiner, F., 2007. Building change detection from Digital Surface Models and multi-spectral images. Photogrammetric Image Analysis, Munich, Germany.

Tian, J., Cui, S., and Reinartz, P., 2014. Building change detection based on satellite stereo imagery and digital surface models. IEEE Transactions on Geoscience and Remote Sensing, 52(1), 406-417.

Zeiler, Matthew D., 2012. ADADELTA: An Adaptive Learning Rate Method. arxiv.org/abs/1212.5701.

Zhang, Z., Vosselman, G., Gerke, M., Persello, C., Tuia, D., Yang, M. Y., 2019. change detection between digital surface models from airborne laser scanning and dense image matching using convolutional neural networks. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., IV-2/W5, 453-460. doi.org/10.5194/isprs-annals-IV-2-W5-453-2019