Efficient Building Inventory Extraction from Satellite Imagery for Megacities

Edmond Yat-Man Lo, En-Kai Lin, Velautham Daksiya, Kuo-Shih Shao, Yi-Rung Chuang, and Tso-Chien Pan

Abstract

Accurate building inventories are essential for city planning and disaster risk management. Traditionally generated via census or selected small surveys, these suffer from data quality and/or resolution. High-resolution satellite imagery with object segmentation provides an effective alternative, readily capturing large extents. This study develops a highly automated building extraction methodology for location-based building exposure data from high (0.5 m) resolution satellite stereo imagery. The development relied on Taipei test areas covering 13.5 km² before application to the megacity of Jakarta. Of the captured Taipei buildings, 48.8% are at one-to-one extraction, improving to 71.9% for larger buildings with total floor area >8000 m², and to 99% when tightly-spaced building clusters are further included. Mean absolute error in extracted footprint area is 16% for these larger buildings. The extraction parameters are tuned for Jakarta buildings using small test areas before covering Jakarta's 643 km² with over 1.247 million buildings extracted.

Introduction

The international disaster event database NatCatSERVICE (2019) reported USD $150 billion economic loss worldwide for the year 2018, arising from 820 natural hazard events, of which Asia accounts for 50%. Reliable estimates of the potential losses need to be developed to support effective risk management of such loss events, particularly those occurring in cities/megacities with consequent large socio-economic impacts. This in turn requires accurate descriptions of the exposures along with the hazard levels faced (Grossi et al. 2005). Such exposure data typically requires spatial maps on assets (e.g., buildings and infrastructures) at risk of covering their location, size, and other characteristics such as vulnerability, as depending on the hazard faced. Exposure maps for individual building location, height, and footprint area, and spanning the entire building inventory are needed for overlaying with hazard event maps in detailed loss assessment such as for floods and earthquakes, the two largest perils by loss magnitude for Asia (NatCatSERVICE 2019). However, such data in Asia and Southeast Asia is generally poor in quality, accessibility, and availability.

A detailed estimation of building inventory traditionally uses census data, conducting of surveys, and/or manual processing of satellite/aerial images. While rich in details at an individual building level, these are often expensive and time consuming processes, implying by necessity either coarseness in spatial resolution or in overall coverage (Figueiredo and Martina 2016; Silva et al. 2015). Automatic building footprint (BFT) extraction from high-resolution satellite and aerial imageries are attractive alternatives in terms of data availability, acquisition cost, and the ability to cover large geographical extents (De Angeli et al. 2016; Gunasekera et al. 2015). However, challenges arise from the close proximity of buildings in dense cities, the diversity of building forms, and the level of differentiation from other background objects (Li et al. 2019). Different approaches are reported (Chen et al. 2018; Gavankar and Ghosh 2018; Li et al. 2019; Ok 2013; Huang and Zhang 2012), with multi-resolution segmentation being the most widely used (Belgiu and Drăguţ 2014; Im et al. 2014). More recent studies involve deep learning applied to semantic segmentation algorithms (Lu et al. 2018; Xu et al. 2018; Sun et al. 2018; Im et al. 2014), though a large scale, city-wide application has yet to be reported.

Besides BFT, building height along with spatial location are also needed for building exposure development. Use of aerial imagery and lidar from low-flying aircraft and/or UAVs for generating Digital Surface Models (DSM) and extracting BFT and height have been reported (Haithcoat et al. 2001; Sahar et al. 2010; Awrangjeb et al. 2010; Yuan 2018; Lu et al. 2018; Xu et al. 2018; Sun et al. 2018). While such imagery offers increased resolution, there are inherent difficulties in securing permission to fly over dense, urban areas. It should also be noted that global commercial technology companies, e.g., Google have developed in-house, proprietary algorithms for extracting BFTs and heights from aerial imagery. There have also been recent, major advances in computer vision, and particularly in image segmentation using Deep Learning (DL) techniques (recent reviews are given in Garcia-Garcia et al. (2017) and Minaee et al. (2021)) applied to deep neural networks (DNN) (the most popular being convolution neural networks). Applications are reported for object detection and classification in both urban and nonurban settings (e.g., Zhang and Zhang 2018; Maltezos et al. 2019; Zhang et al. 2019). The specific works for building classification in urban areas are predominantly based on images from low flying platforms such as airborne laser scanning (ALS) or lidar (Maltezos et al. 2019) or multi-view images (Yu et al. 2021). Building footprint delineation over large areas (Wei et al. 2020) and building reconstruction at Level-of-Detail 1 (LoD-1) for small areas (Zhang and Zhang 2018; Yu et al. 2021) has been reported. Most recently (Gui and Qin 2021) applied DL techniques on very high resolution multi-view satellite images (0.3 to 0.5 m ground sampling distances) and orthophotos to achieve up to LoD-2 level of building reconstruction as demonstrated for small areas (0.5 to 2.25 km²) in three cities. Optional incorporation of the public OpenStreetMap building data further enabled refinement in building orientation. Although DL techniques and DNN have exhibited excellent capability for building extraction and reconstruction, its performance is highly correlated with the size and diversity of labelled training data as appropriate for the city scene. Therefore, building footprint and height reconstruction in complex, diverse, and dense city scenes using DNN remains an active area of research with a variety of data types (e.g., ALS, lidar, multispectral), pre- and post-processing strategies (e.g., Gui and Qin 2021), footprint regularization techniques (e.g., Wei et al. 2020), and fusion of networks (e.g., Bittner et al. 2018)

Edmond Yat-Man Lo, Velautham Daksiya, and Tso-Chien Pan are with the Institute of Catastrophe Risk Management, Nanyang Technological University, Block N1, Level B1b, 50 Nanyang Ave, Singapore 639798 (cymlo@ntu.edu.sg).

Edmond Yat-Man Lo and Tso-Chien Pan are also with the School of Civil and Environmental Engineering, Nanyang Technological University, Block N1, 50 Nanyang Ave, Singapore 639798.

En-Kai Lin, Kuo-Shih Shao, and Yi-Rung Chuang are with Sinotech Engineering Consultants, Inc., No. 171, Sec 5, Nanking E. Rd., Taipei City, Taiwan.

Contributed by Zhenfeng Shao, August 23, 2021 (sent for review April 26, 2022; reviewed by Nan Yang, Min Chen).

Photogrammetric Engineering & Remote Sensing
Vol. 88, No. 10, October 2022, pp. 643–652.
0099-1112/22/643–652

doi: 10.14358/PERS.21-00053R2

proposed, with building footprint and height extraction at whole-city scale yet to be demonstrated. An accuracy-quantified, low-cost, highly-automated, and scalable methodology using readily available high-resolution satellite imagery for extracting location based BFTs and heights as demonstrated here for a megacity have yet to be reported.

In the work here, an efficient, scalable BFT and height extraction methodology is developed and piloted over a 27.9 km^2 area in Taipei, Taiwan and further applied to cover the entire megacity of Jakarta, Indonesia covering an area of 643 km^2. The recently developed rational polynomial coefficient (RPC) stereo processor (RSP) (Qin 2016) for DSM generation, and the commercial *eCognition* software (http://www.ecognition.com/) for image segmentation are applied on high (0.5 m pixel) resolution satellite stereo imagery. Over 1.247 million buildings in Jakarta are extracted. It is further shown that only a small fraction of the extracted BFT require manual adjustment, typically larger buildings as comprising less than 10% of the total building count. This represents huge cost and time savings over traditional methods.

The study areas in Taipei and Jakarta along with data used are presented in the section "Study Areas and Data". A description of the BFT extraction methodology and performance evaluation are in sections "Building Footprint and Height Extraction" and "Performance Evaluation". Results from the Taipei pilot area are presented in the section "Taipei Test Areas", while the application to Jakarta presented in the section "Jakarta City", followed by the Conclusion section.

Study Areas and Data

The Taipei 27.9 km^2 pilot area is located in downtown Taipei (Figure 1a), which spans various building types, including residential, commercial, and public buildings. Two test areas comprising 13.52 km^2 within the pilot area are used for building extraction methodology development and validation. These two test areas represented different building characteristics, with the first test area of 8.43 km^2 being in a newer part of the city, while the second area of 5.09 km^2 is in an older part. The test areas include modern, low- to high-rise buildings, and important landmarks, such as the Taipei Train Station and the supertall 101-storey Taipei 101. Parameters in the developed building extraction methodology are tuned for Jakarta's building using Jakarta's two test areas comprising 9.80 km^2 (Figure 1b), before being applied to the whole city covering 643 km^2. The tuning is needed due to the different building characteristics between Taipei and Jakarta. Jakarta's test areas are chosen from Central (Pusat) Jakarta that covers a variety of residential, commercial, and public buildings, and from north (Utara) Jakarta to further cover industrial buildings. The tallest building included is the 47-storey Menara Astra.

Taipei vector data (TVD) developed by Taipei city government in 2010 is available for both Taipei test areas and served as highly accurate ground reference data for accuracy assessment. The TVD is constructed via stereo-plotting of aerial stereo photos, and is continually updated by Sinotech Engineering Consultants Inc. The data contains detailed (1/1000 scale) boundaries of surface structures including building rooftops in vector format and rooftop height values. Figure 2 (upper panels) show histograms of the 13,208 TVD buildings' BFT areas and heights over the test areas. Only footprint areas >50 m^2 are considered, as Taipei government regulations deem land plots ≤67 m^2 as nonbuildable. The first and second test areas have 6745 and 6459 buildings, respectively, with most having BFT areas ≤2000 m^2.

Similar ground reference data did not exist for Jakarta. Therefore, satellite images are manually stereo-plotted to extract accurate individual building BFT and height for use as ground reference data. The accuracy of the manual extraction was confirmed by two independent polygon extractions done on a small subset of 114 buildings (BFT areas up to 4200 m^2) with the mean absolute difference in extracted BFT areas being <5%. The ground reference set has 11,626 buildings with BFT areas mostly ≤2000 m^2 as in Taipei but with the peak shifted to smaller BFT values (Figure 2, lower panels). Most of the buildings are also lower, ≤3 stories.

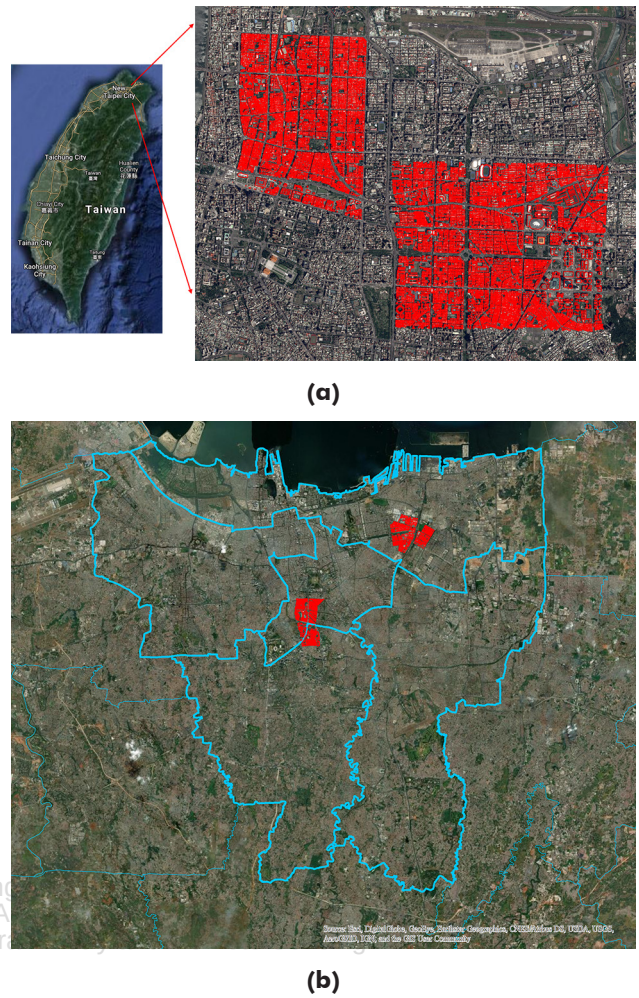


Figure 1. Study areas comprising (a) Taipei pilot area and (b) Jakarta city with its five regions. Test areas are indicated in red.

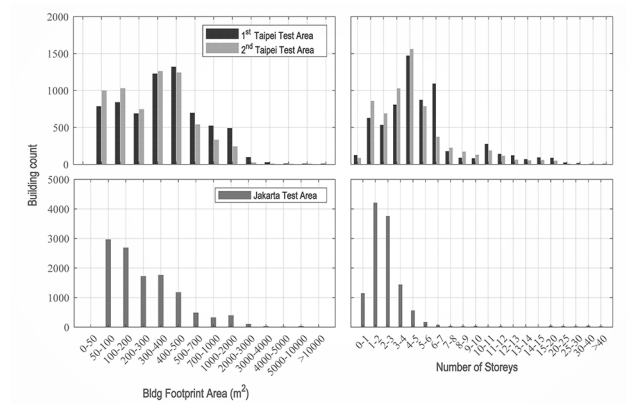


Figure 2. Distribution of building footprint (BFT) area and building height in ground reference data for test areas of Taipei (upper panels) and Jakarta (lower panels). An inter-story height of 3.45 m was used to covert height values to stories.

Regarding sources of satellite stereo images, images from WorldView series, *GeoEye-1*, *QuickBird*, *Pleiades 1A/1B* series, *KOMPSAT-3A/3A*, *CARTOSAT*, and *DMC3*/Triple Sat were evaluated for suitability. The *Pleiades 1A/1B* series were chosen as it allowed tight stereo angle specifications, providing ortho-rectified color data at 0.5 m resolution with high revisit interval of ~two days over Southeast

Asia. A tight stereo angle control was needed to capture tall buildings, while a high revisit rate allowed for cloud/mist free images even in wet, tropical regions, such as that for Jakarta. However, it should be noted that with the rapid advances in commercial satellite imaging, other satellite imagery that meet or exceed these requirements can equally be used.

Building Footprint and Height Extraction

Figure 3 shows a schematic of the BFT extraction methodology. A high degree of automation was achieved via use of RSP, *eCognition* and Geographical Information System (GIS) software, augmented with internally developed software. For DSM generation, the recently available RSP software shown to be particularly suited for use with large scale satellite stereo images (Qin 2016), and the more conventional SOCET GXP (GXP) are assessed with improved results using RSP obtained. In particular, DSM generation in RSP was performed by applying aerial triangulation, referring to ground control points and image matching.

The BFT shape extraction comprise automated segmentation with edge regularization along with a manual adjustment on a small percentage of extracted polygons (Figure 3). The segmentation and regularization procedures followed that of Kuo et al. (2018) and Su et al. (2015). Segmentation was applied only on preprocessed, built-up areas with roads and vegetation surfaces removed. Road surfaces were defined using road vectors from Open Street Map (OSM), with a road buffer width ranging from 1 m (small alleys) to 60 m (arterial road) applied as depending on the road type. Vegetation areas were classified using a Normalized Difference Vegetation Index (NDVI) (Huete et al. 2002), which is a normalized ratio of near infrared (IR) and red bands. Here a calibrated cut-off NDVI value of 0.25 was applied. The *eCognition* software was used for image segmentation on an Object Height Model (OHM), being the difference between DSM and DEM, for high-rise building, and on ortho-rectified images for low-rise buildings, following the segmentation algorithm of Baatz and Schäpe (2000). The first segmentation was exercised on an edge-preserving Kuwahara filtered (sharpened) OHM covering the built-up area following a

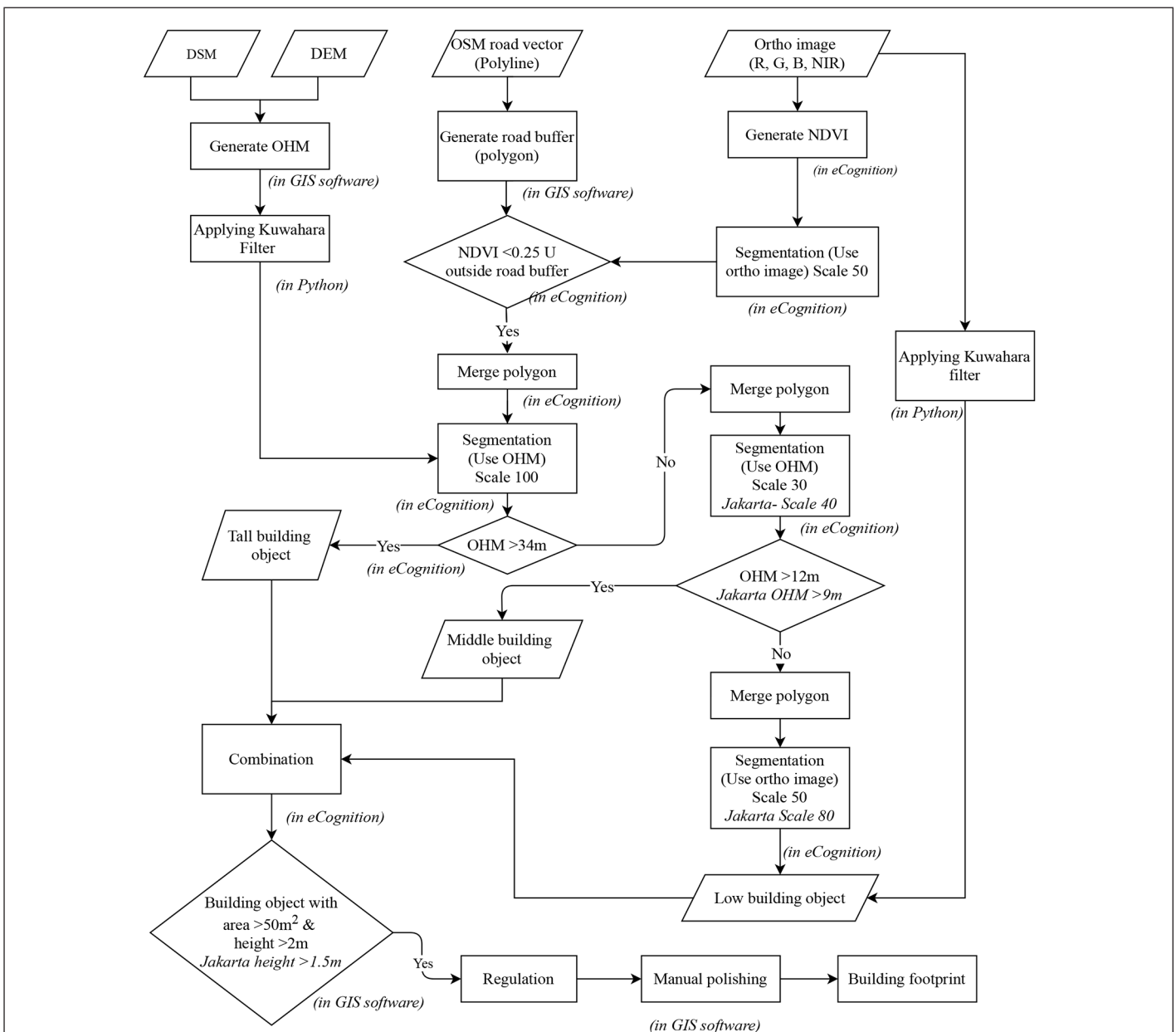


Figure 3. Building footprint (BFT) shape and height extraction methodology flow chart. Parameters, unless separately indicated for Jakarta, are same for Taipei and Jakarta.

building categorization based elevation values (see Figure 3). Regions with high elevations corresponded to larger building objects, which necessitated a larger scale parameter. For low elevation buildings, use of OHM consistently produced over-segmentation with many small objects identified. Rather segmentation on ortho-rectified color images produced satisfactory results for the low elevation buildings. Since the segmentation was based roof top elevations (high- and mid-elevation regions) and roof top features (low elevation region), the segmented objects were found to be sensitive to smaller roof top structures such as lift shafts, water tanks, or in the case of low residential buildings, small (often illegal) roof top additions. As these roof top features did not signify distinct buildings, adjacent polygons in mid- and low-elevation regions are merged if the height difference is less than 1.5 m. Polygons with BFT area ≤50 m^2 (nonbuildable land lots) or height ≤2 m (i.e., nonbuilding objects of low height) were also removed.

The regularization step following Kuo et al. (2018) is to partially correct for jagged lines in the segmented polygon shapes, removing/merging small, extraneous extracted polygons, and straightening of polygon edges. In dense urban areas such as Taipei and Jakarta, many buildings are connected to each other and edges between such adjacent buildings would share the same geometry. These shared edges were simplified and regularized simultaneously to retain the shared edge. A small percentage of the polygons required manual adjustment on footprints by cross-comparison with Google maps and Google Streetview. Manual polishing is conducted on larger size buildings as defined by total floor area (TFA) being ≥8000 m^2 for Taipei and ≥2500 m^2 for Jakarta. This comprised less than 10% total number of buildings and indeed this 10% value was used as a guide in setting the TFA cutoff values for the city-specific larger buildings and thus, the manual effort required. After BFT shape extraction, building height was determined by averaging the OHM values within the extracted polygon shape with allowance of a boundary buffer and removal of outliers. In setting the TFA, an inter-story height of 3.45 m representing an average story height (residential 3.3 m and commercial 3.6 m) in Taipei and Jakarta is used for converting extracted building height to number of stories, which when multiplied by the BFT area gives the TFA.

The extraction methodology development involved a trial-and-error process as guided by a comparison of the extracted building polygons with the ground reference TVD data from Taipei first test area and using Taipei second test area as verification. The parameters of the extraction methodology are expected to have values for Jakarta different from Taipei's due to their different building characteristics (see Figure 2). These are tuned as guided by results from the Jakarta test areas. The notable differences are: <9 m building height is defined as low rise buildings in Jakarta while <12 m is used for Taipei. The segmentation scale for middle and low rise are, respectively, 40 and 80 for Jakarta, while they are 30 and 50 for Taipei. All processing unless otherwise indicated, are done within an ArcGIS environment. The most time intensive step is in the manual polishing performed on the small number of large, extracted building polygons. This required a modest 40 man-weeks (four summer interns were deployed) to cover whole Jakarta spanning 643 km^2.

Performance Evaluation

The detected building polygon areas are first assessed for an overall building area detection performance via standard metrics of True Positive/Negative (TP/TN) and False Positive/Negative (FP/FN). Here TP/TN denote areas correctly classified as spanning buildings/non-building areas, FP denotes area incorrectly classified as building area, and FN as area incorrectly classified as nonbuilding area. From these, quality metrics of quality percentage (QP), detection rate (DR), and overall accuracy (OA) are computed. QP accounts for both boundary delineation accuracy and building detection rate, DR denotes the percentage building area correctly detected, and OA the percentage of building and non-building areas correctly detected.

$$QP = \frac{TP}{TP + FP + FN}; DR = \frac{TP}{TP + FN}; OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

However, beyond such metrics on overall building area being accurately detected, the individually extracted buildings (or polygons) do not necessarily have one-to-one match with the actual, with a common challenge of closely spaced buildings being extracted as one building in dense urban settings. This necessitated a more detailed object-level accuracy assessment, as achieved here by categorizing the extracted polygons into five cases below:

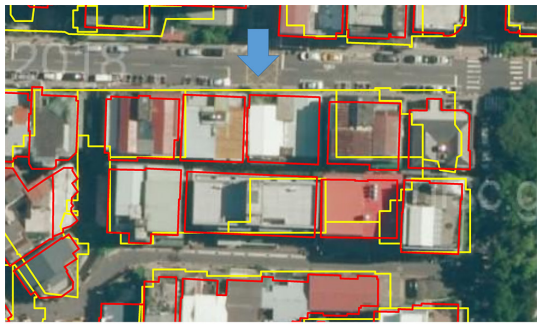
- Case 1 (extraneous): An extraneous polygon when the centroid of the extracted polygon does not fall within the boundary of any ground reference polygon.
- Case 2 (one-to-one): One-to-one (desired) match between an extracted polygon and a ground reference polygon. Here the centroids of extracted polygon and ground reference polygon fall within the boundaries of each other.
- Case 3 (one-to-many): The extracted polygon enclosed several ground reference polygons as defined by the centroids of the ground reference polygons falling within the boundary of the extracted polygon.
- Case 4 (many-to-one): Multiple polygons are extracted from one single ground reference polygon, and the boundary of the ground reference polygon encompasses the centroids of several extracted polygons.
- Case 5 (missed): Polygon missed in extraction.

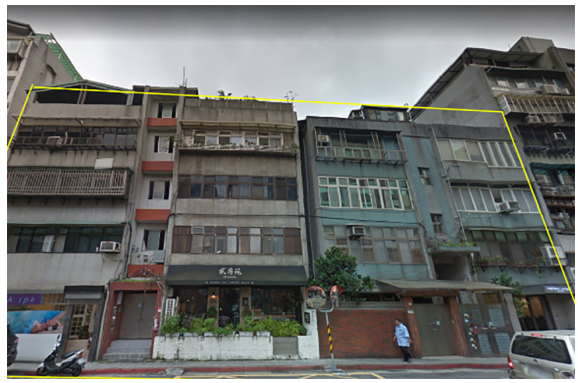
Case 3 occurs when a cluster of tightly spaced ground reference buildings of similar height and inseparable from aerial views are grouped together in the extracted polygon (see Figure 4a). This often happens in dense urban settings with a prevalence of rows of tightly-spaced, inseparable buildings. Case 4 corresponds to buildings with complex roof top configurations or features resulting in multiple polygons being extracted within a single ground reference polygon (Figure 4b). As such, a count accuracy analysis across the five cases is first performed before an accuracy evaluation on the values of extracted BFT area and height. The analysis is performed separately for Taipei first test area (used for methodology development, i.e., a calibration) and tested independently over the second test area (i.e., a verification).

Results and Discussion

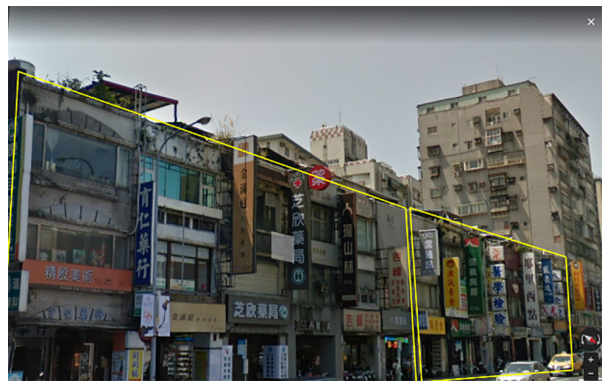
DSM generation using RSP and the more conventional GXP were first quantitatively compared for the Taipei test areas. Building rooftop elevations from TVD (subset of 125 buildings) are used to determine convergence angle requirements in the satellite stereo imagery needed for capturing tall buildings. The 125 TVD elevation values ranged up to 174 m, with a further value at 391 m corresponding to the super tall Taipei 101. Building rooftop elevation values calculated using RSP are more accurate than using GXP with regions of higher elevation having better delineation. Over 92% of the 125 buildings have elevation error <3 m at 14° convergence angle, this improving to 94% at 10°. The sole exception is Taipei 101, where RSP significantly underestimated and GXP even more so. The subsequent images used below for Jakarta are at 12° convergence angle. For OHM generation, a DEM was derived by sampling the RSP generated DSM ground points in free space such as parks and roads. A total of 432 ground points within and around the Taipei test areas are interpolated for DEM generation.

Figure 5 shows the extracted BFT polygons for the Taipei and Jakarta test areas. QP, DR, and OA values for Taipei first test area are 71.4%, 83.9%, and 87.5%, respectively, and 73.6%, 85.6%, and 86.8%, respectively, for the second test area. These values are within reported ranges for building detection applications from satellite imagery (Hermosilla et al. 2011; Ghandour and Jezzini 2018; Jin and Davis 2005; Lee et al. 2003), though higher ranges are reported by Khoshelham et al. (2010) who used multi-source (including lidar) data. The QP, DR, and OA values are 79.3%, 88.3%, and 91.4%, for the Jakarta test areas, slightly better than Taipei's.



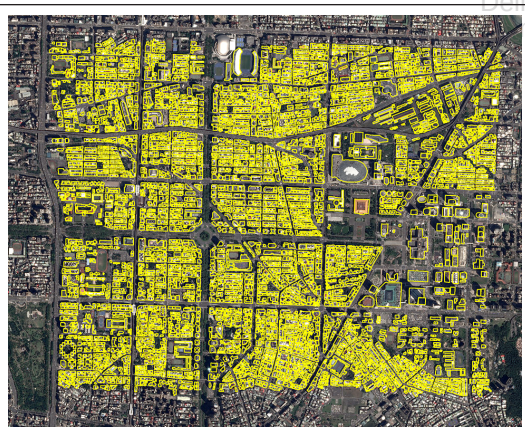


(a)

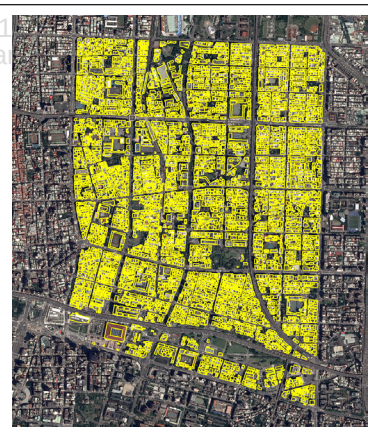


(b)

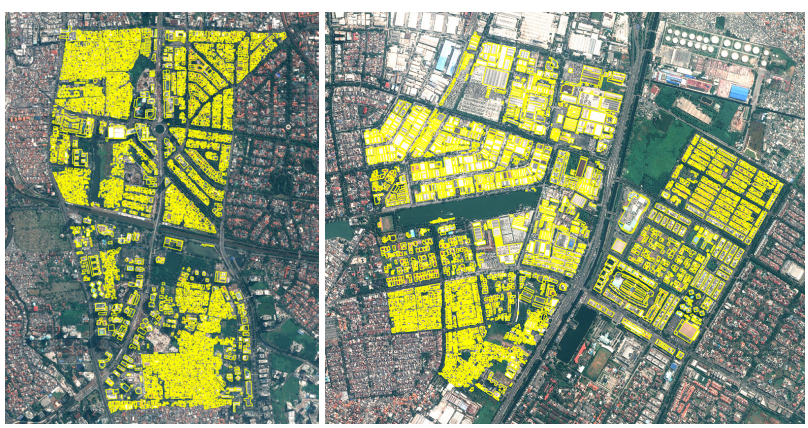
Figure 4. Buildings categorized as (a) Case 3: unseparated buildings due to small separation and small difference in roof height and (b) Case 4: over-segmented due to various rooftop structures. Examples shown are from Taipei test areas.



(a) Extracted BFT - Taipei 1st test area



(b) Extracted BFT - Taipei 2nd test area



(c) Extracted BFT – Jakarta test areas

Figure 5. Samples of extracted building footprint (BFT) from Taipei and Jakarta test areas.

Taipei Test Areas

The number of building polygons falling in each Case 1 to 5 are shown in Table 1 for Taipei's two test areas individually and combined. Focusing on the first test (calibration) area, 46.9% of the extracted polygons have a one-to–one (Case 2, ideal) extraction, while Cases 2 and 3 combined totals 67.0%. For estimating total building value exposed to hazards, Cases 3 and 4 as representing dense building clusters would also be considered as correct extraction. This is because Case 3 has one extracted building (polygon) encompassing several tightly-spaced, i.e., essentially inseparable buildings that would be very similar in height and in structural characteristics, and therefore in building exposure and vulnerability characteristics. This also applies for Case 4 where several extracted building polygons with very similar characteristics collectively represent one actual building. With this, the combined polygons over Cases 2, 3, and 4 comprise 95.07% of the extracted. Only a small percentage at 4.93% are in Case 1 (extraneous) polygons. It is noted these percentage counts are based on the total extracted polygon count (i.e., Sum Cases 1 to 4). The number of missed polygons (Case 5) is 713, i.e. 13.5% relative to the total number of extracted polygons. Similar extraction performance is seen for Taipei second test (verification) area. The extracted have 50.7% one-on–one (Case 2), and 95.3% in combined Cases 2, 3, and 4, with only a small 4.71% Case 1 (extraneous polygon). The number of missed polygons (Case 5) is 440, i.e., 8.5% relative to the total number extracted.

The extraction performance, particularly Case 2, is improved for the larger TFA building polygons (Table 1). For building TFA >4000 m^2, Case 2 extraction is slightly improved at 51.8% from 48.8% for the combined Taipei test areas and improving significantly to 71.9% for TFA >8000 m^2. Thus, progressively more of the larger buildings, and thus building values exposed to hazards, are extracted on one-to-one basis. The percentage in extracted Cases 3 and 4 (i.e., building clusters) is essentially unchanged for TFA >4000 m^2, but notably reduced for TFA >8000 m^2 at 27.4%. Greater than 99% accuracy over combined Cases 2 to 4 is achieved for the larger buildings with TFA >4000 m^2 and >8000 m^2 (Table 1); this also holds for the individual test areas (not shown).

The percentage distribution of building polygon counts across Cases 1–5 and cumulative for Taipei's combined test areas are plotted in TFA bins (Figure 6). It is evident that Case 2 (one-to-one) extraction (red bars) becomes dominant at larger building TFA bins. In contrast, Case 1's 506 extraneous polygons (Table 1) and Case 5's 1153 missed polygons were mostly confined to buildings with small TFA of <1000 m^2. This again indicates that the larger buildings are better extracted. Successfully capturing such large buildings is key towards capturing the city's entire built-up TFA, and thus a city's inventory of building values and exposure. Thus Figure 6 shows that the larger TFA>8000 m^2 buildings contributing 43% of the test areas' cumulative TFA.

Detailed error analysis on extracted BFT area and height values are next discussed considering Cases 2, 3, and 4 individually, and covering Taipei first (calibration) and second (verification) test area separately. Table 2 summarizes the BFT area error, dA, for all buildings and building with large TFA >4000 m^2 and >8000 m^2. The results show that 67–74% of all buildings in Cases 2 and 3 had error abs(dA) <30% across the two test areas. In this error calculation, Case 3 used the total BFT areas of the encompassed ground truth (TVD) buildings as these are corresponded to tightly-spaced, inseparable buildings. As expected, there is a significant improvement for the larger TFA buildings where 71–89% in Cases 2 and 3 has abs(dA) <30% for building with TFA >4000 m^2, and >92% for TFA >8000 m^2. By comparison, the errors for

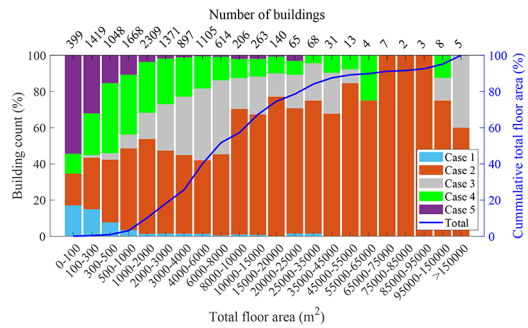


Figure 6. Building count (%) over Cases 1–5 with buildings grouped in total floor area (TFA) bins for combined Taipei test areas. Upper horizontal axis indicates number of buildings in each TFA bin. Solid blue line and right vertical axis indicates the cumulative TFA (%) spanned over the test areas.

Table 1. Count Analysis across Cases 1 to 5 building polygons for Taipei and Jakarta test areas. Percentage numbers in parentheses are based on number of extracted polygons (sum Cases 1 to 4).

	Taipei					Jakarta Test Area		
	First Test Area (Calibration)	Second Test Area (Verification)	Combined Test Areas					
Case			All	TFA >4000 m^2	TFA >8000 m^2	All	TFA >4000 m^2	TFA >8000 m^2
1 (extraneous), n (%)	261 (4.93)	245 (4.71)	506 (4.82)	24 (0.96)	6 (0.75)	642 (6.49)	3 (0.53)	1 (0.28)
2 (one-to-one), n (%)	2485 (46.9)	2635 (50.7)	5120 (48.8)	1301 (51.8)	577 (71.9)	4851 (49.0)	481 (85.1)	299 (84.2)
3 (one to many), n (%)	1065 (20.1)	928 (17.9)	1993 (19.0)	834 (33.2)	145 (18.1)	1793 (18.1)	61 (10.8)	46 (13.0)
4 (many to one), n (%)	1483 (28.0)	1390 (26.7)	2873 (27.4)	352 (14.0)	75 (9.3)	2606 (26.3)	20 (3.5)	9 (2.5)
5 (missed)	[713]	[440]	[1153]	[23]	[12]	[1043]	[8]	[1]
Total Extracted (Sum Cases 1 to 4)	5294	5198	10,492	2511	803	9892	565	355
Sum Cases 2 to 4/Total Extracted (%)	95	95.3	95.2	99.0	99.3	93.4	99.5	99.7

TFA = total floor area.

Table 2. Count and percentage of buildings in Taipei test areas with building footprint error abs(dA) < 20% and < 30%.

		Taipei First Test Area						Taipei Second Test Area					
		All buildings		TFA >4000 m^2		TFA >8000 m^2		All buildings		TFA >4000 m^2		TFA >8000 m^2	
Case		Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Case 2	abs(dA) < 20%	1345/2485	54	600/783	77	330/376	88	1493/2635	57	376/518	73	160/201	80
	abs(dA) < 30%	1672/2485	67	678/783	87	356/376	95	1864/2635	71	459/518	89	188/201	94
Case 3	abs(dA) < 20%	580/1065	54	274/505	54	92/118	78	514/928	55	185/329	56	14/27	89
	abs(dA) < 30%	759/1065	71	359/505	71	108/118	92	685/928	74	244/329	74	27/27	100
Case 4	abs(dA) < 20%	121/1483	8	47/230	20	8/40	20	114/1390	8	23/122	19	6/35	17
	abs(dA) < 30%	204/1483	14	69/230	30	11/40	28	176/1390	13	36/122	30	7/35	20

TFA = total floor area.

Case 4 buildings are much larger, as expected, since the one extracted polygon encompassed several TVD polygons, but only one TVD polygon was matched with the extracted for the dA error calculation.

Table 3 shows the errors dH in the extracted building heights. For Taipei first (calibration) test area, 91% and 94% of all buildings in Case 2 and 3, respectively had abs(dH) <3 m and with the accuracy being essentially unchanged for the larger TFA buildings, whilst 81% of all buildings in Case 4 had abs(dH) <3 m. The height accuracy for the second (verification) test area was less by comparison, which was attributed to the TVD height values being less accurate due to this area being in an older part of Taipei with older TVD.

The Mean Absolute Errors (MAE) for BFT area, height, and TFA are listed in Table 4 for the combined Taipei test areas. Results at four TFA ranges are shown, comprising all TFA values, TFA ≤4000 m^2, TFA between 4000 m^2–8000 m^2, and TFA >8000 m^2. While the MAE values on both BFT area and TFA over all buildings (all TFA values) are high at 37%–40%, the values are much reduced for buildings with TFA between 4000–8000 m^2 and TFA >8,000 m^2, with the MAE in TFA being at 29% and 17%, respectively.

The scalability of the building extraction algorithm is next demonstrated by applying the extraction framework to the full Taipei pilot area of 27.9 km^2 (see Figure 7a). In total, 20,597 building polygons are extracted, of which 2355 are with large BFT area (>1000 m^2) and 1957 have height >10 floors. Also 47% (i.e., almost half) of the entire TFA over the pilot area is contributed by the larger buildings (TFA >8000 m^2); these buildings comprise only 9% of total building count. Furthermore, these buildings are expected to have small MAE as their extracted BFT area, height, and TFA (Table 4).

Jakarta City

The full scalability of the building extraction is demonstrated for the megacity of Jakarta covering an area of 643 km^2. The extraction algorithm parameters are first tuned using the Jakarta test areas' ground reference to account for building characteristics being different from Taipei's. The earlier Table 1 also shows the number of extracted building polygons from Jakarta test areas as falling into Cases 1 to 5. Compared to the Taipei test areas, the algorithm had comparable performance in Case 2 (one-to-one) extraction, and with a better improvement for the larger TFA buildings at 84%–85% extraction, with the percentages for Cases 3 and 4 correspondingly reduced. Case 1 extraneous and Case 5 missed polygons remain small as with Taipei's test areas. Figure 8 plots the percentage distribution of building polygons counts across Cases 1–5 and cumulative for Jakarta's combined test areas against TFA bins (i.e., as shown earlier in Figure 6 for Taipei test areas). As with the Taipei test area results, Case 2 (one-to-one) extraction becomes dominant at large TFA bins, while Case 1 (extraneous) polygons and Case 5 missed polygons are mostly confined to small TFA bins. Similarly, the larger building contributed disproportionally to the cumulative TFA, e.g., building with TFA >4000 m^2 contributed 70.5% of Jakarta's test areas cumulative TFA.

In terms of dA errors on BFT area, 52.4% of all buildings are extracted with abs(dA) < 30%, and further improved for larger TFA (>4000 m^2) buildings at 92.9%. The error in building height dH has 75.7% of extracted buildings having abs(dH) < 3 m, and 75.4% (i.e., essentially unchanged) for the larger TFA (>4000 m^2) buildings. The MAE values (Table 4) are comparable to that achieved for Taipei test areas, and better for the larger TFA buildings.

The algorithm extracted over 1.247 million buildings for the megacity of Jakarta (Figure 7b). Extracted BFT area, height, and TFA statistics are shown in Figure 9. Jakarta's buildings largely are low rise with 59% (32%) of the buildings are at one (two) story (Figure 9a), and only 1744 buildings having >10 stories. 78% of the buildings have BFT areas of 50–300 m^2, while only 13,302 buildings have BFT >1000 m^2 area (Figure 9b). Small TFA <4000 m^2 buildings contributes to 76% of Jakarta's entire TFA (Figure 9c), this representing 99% of the total number of buildings, whilst large buildings of TFA >8000 m^2 contributes 18% of Jakarta's entire TFA, i.e., a smaller percentage when compared to the Jakarta (also Taipei) test areas. This is because these test areas, being in their respective downtown core, are more populated with high-rise and large footprint buildings.

Conclusion

The work demonstrates a highly efficient and automated BFT and height extraction methodology using high-resolution satellite stereo images and off-the-shelf software, with an application to the megacity of Jakarta. The methodology is developed using small Taipei test areas where accurate ground reference TVD is available. Differences in building characteristics between the two cities are accounted for via tuning of algorithm parameters using small Jakarta test areas.

The results on extracted buildings are analyzed over the test areas, both on the extraction count over buildings, which can be closely

Table 3. Count and percentage of buildings in Taipei test areas with height error abs(dH) < 2 m and < 3 m.

		Taipei First Test Area						Taipei Second Test Area					
		All buildings		TFA >4000 m^2		TFA >8000 m^2		All buildings		TFA >4000 m^2		TFA >8000 m^2	
Case		Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Case 2	abs(dH) < 2 m	2113/2485	85	675/783	86	316/376	84	1703/2635	65	312/518	60	101/201	50
	abs(dH) < 3 m	2263/2485	91	699/783	89	331/376	88	2116/2635	80	371/518	72	126/201	63
Case 3	abs(dH) < 2 m	959/1065	90	457/505	90	99/118	84	729/928	79	269/329	82	23/27	85
	abs(dH) < 3 m	1001/1065	94	473/505	94	104/118	88	832/928	90	299/329	91	24/27	89
Case 4	abs(dH) < 2 m	1045/1483	70	187/230	81	26/40	65	817/1390	59	62/122	51	14/35	40
	abs(dH) < 3 m	1201/1483	81	199/230	87	28/40	70	996/1390	72	76/122	62	18/35	51

TFA = total floor area.

Table 4. Mean Absolute Errors (MAE) (%) in building footprint (BFA), height, and total floor area (TFA) for buildings in the combined test areas of Taipei.

	Taipei				Jakarta			
Parameter	All TFA buildings	TFA ≤4000 m^2	TFA: 4000–8000 m^2	TFA >8000 m^2	All TFA buildings	TFA ≤4000 m^2	TFA: 4000–8000 m^2	TFA >8000 m^2
BFT MAE	37	42	27	16	37	39	9	8
Height MAE	12	14	8	7	28	29	14	8
TFA MAE	40	45	29	17	48	49	20	15

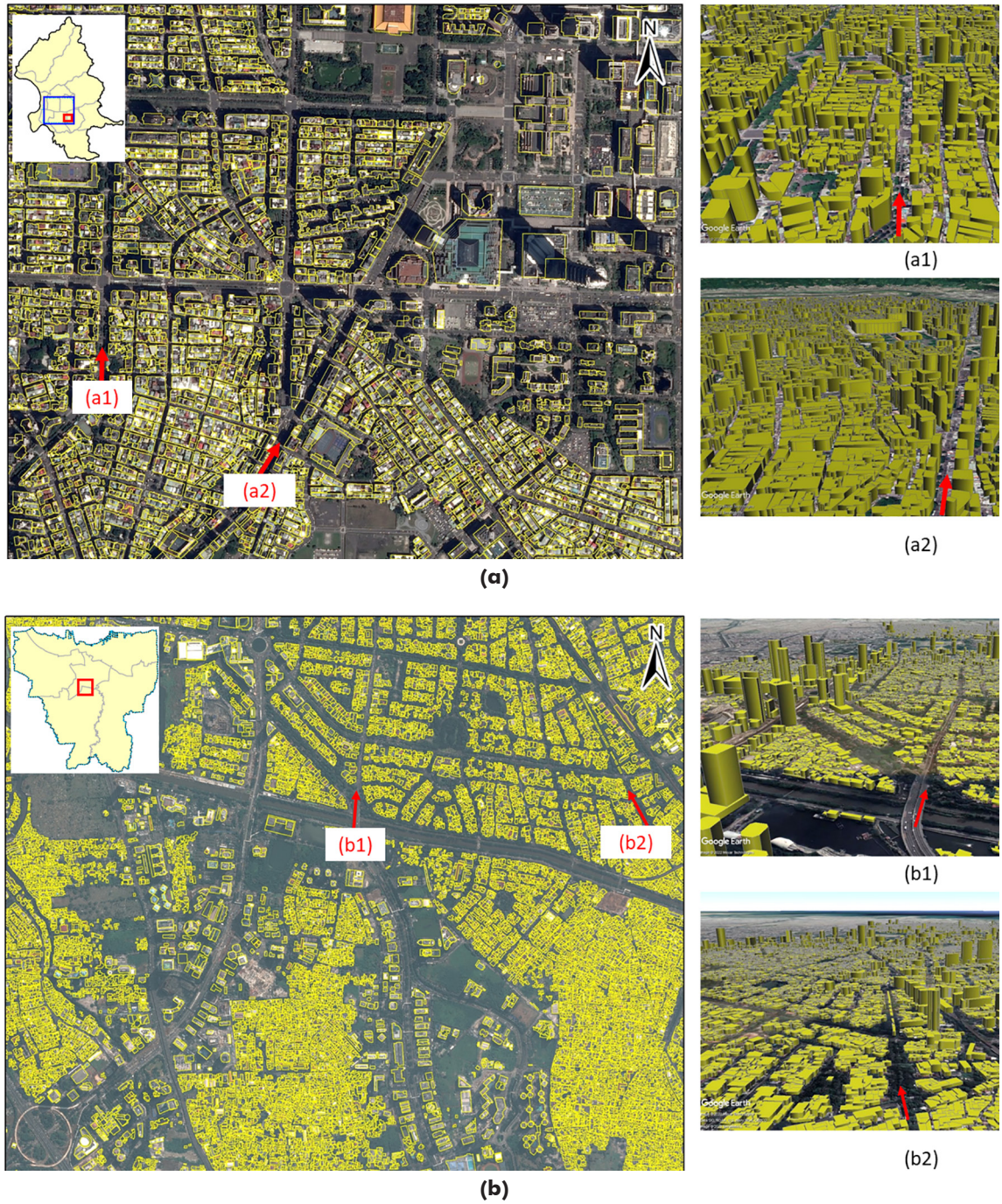


Figure 7. Building polygons extracted from (a) Taipei 27.9 km^2 pilot area and (b) Jakarta over 643 km^2. Solid blue lines in upper left inset indicate the full areal extent covered, and solid red lines show the areal extent of an expanded two-dimensional view with further detailed three-dimensional views shown in the left panels at locations indicated by the red arrows.

spaced in dense urban areas, and on errors in the extracted BFT area, height, and TFA. It is shown that buildings captured in Taipei test areas are at one-to-one (Case 2) extraction for 48.8% of the captured buildings, improving to 71.9% for larger buildings with TFA >8000 m^2, and further reaches 99% accuracy when closely-spaced, inseparable building clusters of similar height and structural characteristics are included, as appropriate for building exposure development. Extraneous and missed building were small in number, and notably are largely confined to small TFA buildings. The MAE in BFT area, while being at 37% over all captured buildings, reduced significantly to 27% for larger TFA buildings (TFA between 4000–8000 m^2), and to 16% for buildings with TFA >8000 m^2. Similar extraction performance and accuracies hold for the Jakarta test areas. It is also shown that the larger TFA buildings, while small in count number, contributes disproportionally to the cumulative TFA for both Taipei and Jakarta test areas, e.g., Jakarta's test areas located in the downtown core have buildings with TFA >4000 m^2 contributing to 70.5% of the cumulative TFA, whilst over the entire Jakarta of 643 km^2, such buildings with >4000 m^2 TFA contributes 24% of the cumulative TFA.

The analyses indicate that the extraction methodology is effective, even for megacities, accurately capturing building inventory covering areas and heights via readily available satellite imagery. This provides

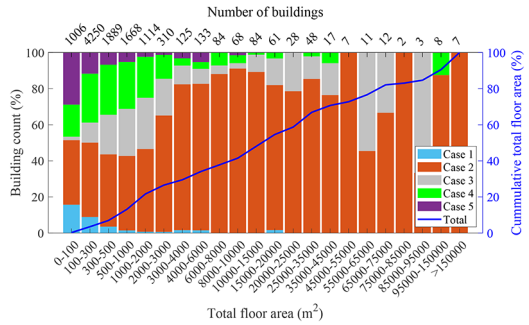


Figure 8. Building count (%) over Cases 1–5 with buildings grouped in total floor area (TFA) bins for combined Jakarta test areas. Upper horizontal axis indicates number of buildings in each TFA bin. Solid blue line and right vertical axis indicates the cumulative TFA (%) spanned over the test areas.

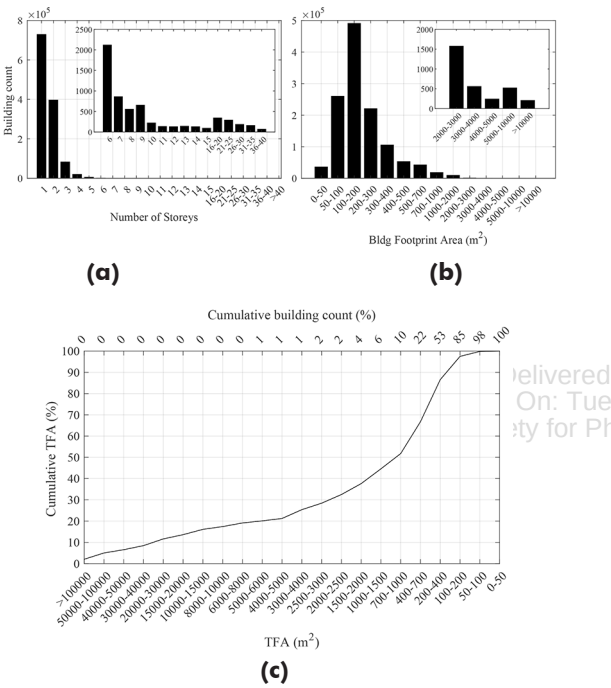


Figure 9. Statistics on buildings in Jakarta. Note in (c) total floor area (TFA) bins are plotted in decreasing bin size and the upper horizontal axis indicate cumulative building count (%).

a cost-effective, readily deployable option for quantifying city-wide building inventory needed in city planning, and for risk analysis under hazards. It would therefore find use in national and regional government units involved in disaster planning and management. Lastly, the extraction methodology can readily take advantage of increasingly more accurate satellite imagery or aerial/UAV images as they become available.

References

Awrangjeb, M., M. Ravanbakhsh and C. S. Fraser. 2010. Automatic detection of residential buildings using LIDAR data and multispectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(5):457–467.

Baatz, M. and A. Schäpe. 2000. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. Pages 12–23 in *Proceedings Angewandte Geographische Informationsverarbeitung XII*, *Beiträge zum AGIT-Symposium Salzburg 2000*. Edited by J. Strobl. Karlsruhe, Germany: Herbert Wichmann Verlag.

Belgiu, M. and L. Drăguţ. 2014. Comparing supervised and unsupervised multiresolution segmentation approaches for extracting buildings from very high resolution imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96:67–75.

Bittner, K., F. Adam, S. Cui, M. Körner and P. Reinartz. 2018. Building footprint extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(8):2615–2629.

Chen, R., X. Li and J. Li. 2018. Object-based features for house detection from RGB high-resolution images. *Remote Sensing* 10(3):451.

De Angeli, S., F. Dell'Acqua and E. Trasforini. 2016. Application of an earth-observation-based building exposure mapping tool for flood damage assessment. In *Proceedings 3rd European Conference on Flood Risk Management (FLOODrisk 2016)*. Edited by M. Lang, F. Klijn and P. Samuels. *EDP Science*. https://doi.org/10.1051/e3sconf/20160705001.

Figueiredo, R. and M. Martina. 2016. Using open building data in the development of exposure data sets for catastrophe risk modelling. *Natural Hazards and Earth System Sciences* 16(2):417–429. https://doi.org/10.5194/nhess-16-417-2016.

Garcia-Garcia, A., S. Orts, S. Oprea, V. Villena-Martinez and J. G. Rodríguez. 2017. A review on deep learning techniques applied to semantic segmentation. arXiv:1704.06857v1.

Gavankar, N. L. and S. K. Ghosh. 2018. Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing* 51(1):182–193. https://doi.org/10.1080/22797254.2017.1416676.

Gui, S. and R. Qin. 2021. Automated LoD-2 model reconstruction from very-high-resolution satellite-derived digital surface model and orthophoto. *ISPRS Journal of Photogrammetry and Remote Sensing* 181:1–19.

Ghandour, A. J. and A. A. Jezzini. 2018. Autonomous building detection using edge properties and image color invariants. *Buildings* 8(5):65.

Grossi, P., H. Kunreuther and D. Windeler. 2005. An introduction to catastrophe models and insurance. In *Catastrophe Modeling: A New Approach to Managing Risk*, edited by P. Grossi and H. Kunreuther, 23–42. Boston, Massachusetts: Springer.

Gunasekera, R., O. Ishizawa, C. Aubrecht, B. Blankespoor, S. Murray, A. Pomonis and J. Daniell. 2015. Developing an adaptive global exposure model to support the generation of country disaster risk profiles. *Earth-Science Reviews* 150:594–608. https://doi.org/10.1016/j.earscirev.2015.08.012.

Haithcoat, T. L., W. Song and J. D. Hipple. 2001. Building footprint extraction and 3-D reconstruction from LIDAR data. In *Proceedings IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas (Cat. No.01EX482)*, held in Rome, Italy, 8–9 November 2001. https://doi.org/10.1109/DFUA.2001.985730.

Hermosilla, T., L. A. Ruiz, J. A. Recio and J. Estornell. 2011. Evaluation of automatic building detection approaches combining high resolution images and lidar data. *Remote Sensing* 3(6):1188–1210.

Huang, X. and L. Zhang. 2012. Morphological building/shadow index for building extraction from high-resolution imagery over urban areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5(1):161–172. https://doi.org/10.1109/JSTARS.2011.2168195.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao and L. G. Ferreira. 2002 Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83(1):195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

Im, J., L. J. Quackenbush, M. Li and F. Fang. 2014. Optimum scale in object-based image analysis. In *Scale Issues in Remote Sensing*, edited by Q. Weng. Hoboken, New Jersey: Wiley. https://doi.org/10.1002/9781118801628.ch10.

Jin, X. and C. H. Davis. 2005. Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Advances in Signal Processing* (14):745309. https://doi.org/10.1155/ASP.2005.2196.

Khoshelham, K., C. Nardinocchi, E. Frontoni, A. Mancini and P. Zingaretti. 2010. Performance evaluation of automated approaches to building detection in multi-source aerial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(1):123–133. https://doi.org/10.1016/j.isprsjprs.2009.09.005.

Kuo, H-Y., J-Y. Rau and EY-M. Lo. 2018. Building footprint regulation assisted by road vector derived from open street map. *Proceedings International Symposium on Remote Sensing*, held in Pyeongchang, Korea, 9–11 May 2018.

Lee, D. S., J. Shan and J. S. Bethel. 2003. Class-guided building extraction from Ikonos imagery. *Photogrammetric Engineering & Remote Sensing* 69(2):143–150. https://doi.org/10.14358/PERS.69.2.143.

Li, W., C. He, J. Fang, J. Zheng, H. Fu and L. Yu. 2019. Semantic segmentation-based building footprint extraction using very high-resolution satellite images and multi-source GIS data. *Remote Sensing* 11(4):403.

Lu, T., D. Ming, X. Lin, Z. Hong, X. Bai and J. Fang. 2018. Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network. *Remote Sensing* 10(9):1496.

Maltezos, E., A. Doulamis, N. Doulamis and C. Ioannidis. 2019. Building extraction from LiDAR data applying deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters* 16(1):155–159.

Minaee, S., Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz and D. Terzopoulos. 2021. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2021.3059968.

NatCatSERVICE–Munich Re. 2019. Münchener Rückversicherungs-Gesellschaft (Munich Reinsurance Company). <https://natcatservice.munichre.com> Accessed 26 March 2021.

Ok, A. O. 2013. Automated detection of buildings from single VHR multispectral images using shadow information and graph cuts. *ISPRS Journal of Photogrammetry and Remote Sensing* 86:21–40. https://doi.org/10.1016/j.isprsjprs.2013.09.004.

Qin, R. 2016. RPC stereo processor (RSP)—A software package for digital surface model and orthophoto generation from satellite stereo imagery. Pages 77–82 in *Proceedings ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, held in Prague, Czech Republic, 12–19 July 2016. Copernicus Publications. https://doi.org/10.5194/isprs-annals-III-1-77-2016.

Sahar, L., S. Muthukumar and S. P. French. 2010. Using aerial imagery and GIS in automated building footprint extraction and shape recognition for earthquake risk assessment of urban inventories. *IEEE Transactions on Geoscience and Remote Sensing* 48(9):3511–3520. https://doi.org/10.1109/TGRS.2010.2047260.

Silva, V., H. Crowley, H. Varum and R. Pinho. 2015. Seismic risk assessment for mainland Portugal. *Bulletin of Earthquake Engineering* 13:429–457. https://doi.org/10.1007/s10518-014-9630-0.

Su, G., W. Qi, S. Zhang, T. Sim, X. Liu, R. Sun, L. Sun and Y. Jin. 2015. An integrated method combining remote sensing data and local knowledge for the large-scale estimation of seismic loss risks to buildings in the context of rapid socioeconomic growth: A case study in Tangshan, China. *Remote Sensing* 7(3):2543.

Sun, Y., X. Zhang, X. Zhao and Q. Xin. 2018. Extracting building boundaries from high resolution optical images and LiDAR data by integrating the convolutional neural network and the active contour model. *Remote Sensing* 10(9):1459.

Wei, S., S. Ji and M. Lu. 2020. Toward automatic building footprint delineation from aerial images using CNN and regularization. *IEEE Transactions on Geoscience and Remote Sensing* 58(3):2178–2189.

Xu, Y., L. Wu, Z. Xie and Z. Chen. 2018. Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing* 10(1):144.

Yu, D., S. Ji, J. Liu and S. Wei. 2021. Automatic 3D building reconstruction from multi-view aerial images with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 171:155–170.

Yuan, J. 2018. Learning building extraction in aerial scenes with convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(11):2793–2798. https://doi.org/10.1109/TPAMI.2017.2750680.

Zhang, L., Z. Shao, J. Liu and Q. Cheng. 2019. Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data. *Remote Sensing* 11(12). https://doi.org/10.3390/rs11121459.

Zhang, L. and L. Zhang. 2018. Deep learning-based classification and reconstruction of residential scenes from large-scale point clouds. *IEEE Transactions on Geoscience and Remote Sensing* 56(4):1887–1897.

IN-PRESS ARTICLES

Sunil Kumar Pundir, Rahul Dev Garg. Development of technique for Vehicle specific off-road trafficability assessment using soil cone index, water index & geospatial data.

Weidong Li, Fanqian Meng, Yongbo Yu. The use of indices and modified U-Net network in improving the classification of planting structures.

Weiwei Liu. Managing earth hazards using deep reinforcement learning algorithm for industrial IoT network.

Itiya Aneece, Prasad S. Thenkabail. New generation hyperspectral sensors DESIS and PRISMA provide improved agricultural crop classifications.

Yanzhe Shi, Yumin Tan, Yunxin Li, Bo Xu. Automatic registration method of multi-source point clouds based on building facades matching in urban scenes.

Zhang Yan, Wang Xiangyu, Zhang Zhongwei, Sun Yemei, and Liu Shudong. Foreground-Aware Refinement Network for Building Extraction from Remote Sensing Images.

Yumin Tan, Yanzhe Shi, Yunxin Li, and Bo Xu. Automatic registration method of multi-source point clouds based on building facades matching in urban scenes.

Yan Lv, Hongwei Guo, Shuanggen Jin, Lu Wang, Haiyi Bian, and Haijian Liu. Permanganate index variations and factors in Hongze Lake from Landsat-8 images based on machine learning.

Songjing Guo, Xueling Wu, Ruiqing Niu, and Wenfu Wu. Exploring spatiotemporal variations and driving factors of urban comprehensive carrying capacity in the Yangtze River Delta urban agglomeration.