Multi-scale solution for building extraction from LiDAR and image data

T. Thuy Vu *
Geoinformatics, Royal Institute of Technology (KTH)
Drottning Kristinas väg 30, 100 44, Stockholm, Sweden
Email: thuyvu@infra.kth.se
* Corresponding author
TEL. 46-8-790-7341
FAX 46-8-790-8580

Fumio Yamazaki
Department of Urban Environment Systems,
Chiba University
1-33 Yayoi-cho, Inage-ku, Chiba, 263-8522, Japan
Email: yamazaki@tu.chiba-u.ac.jp

Masashi Matsuoka
GEO Grid Research Group, Information Technology Research Institute
National Institute of Advanced Industrial Science and Technology (AIST)
Umezono 1-1-1, Tsukuba 305-8568, Japan
Email: m.matsuoka@aist.go.jp

Abstract. This paper presents a multi-scale solution based on mathematical morphology for extracting the building features from remotely sensed elevation and spectral data. Elevation data are used as the primary data to delineate the structural information and are firstly represented on a morphological scale-space. The behaviors of elevation clusters across the scale-space are the cues for feature extraction. As a result, a complex structure can be extracted as a multi-part object in which each part is represented on a scale depending on its size. The building footprint is represented by the boundary of the largest part. Other object attributes include the area, height or number of stories. The spectral data is used as an additional source to remove vegetation and possibly classify the building roof material. Finally, the results can be stored in a multi-scale database introduced in this paper. The proposed solution is demonstrated using the data derived from a LiDAR (Light Detection And Ranging) surveying flight over Tokyo, Japan. The results show a reasonable match with reference data and prove the capability of the proposed approach in accommodation of diverse building shapes. Higher density LiDAR is expected to produce better accuracy in extraction, and more spectral sources are necessary for further classification of building roof material. It is also recommended that parallel processing should be implemented to reduce the computation time.

Keywords: image analysis, mathematical morphology, information storage, remote sensing.

Acknowledgments

The authors would like to thank Mr. Keita Kato of Chiba University for preparing the reference data and performing some accuracy assessment and Dr. Kouichi Hasegawa of Earthquake Disaster Mitigation Research Center for providing the collection of the roof material survey.

1 INTRODUCTION

Building inventory databases provide essential information for earthquake damage assessment. Updating them may benefit from the extraction of building features from remotely sensed data, which has been progressively developed. Recent developments in feature extraction image analysis show the following typical tendencies: there has been a transition to 3D, object-oriented, hierarchical and multi-scale approaches are often used, and more attention has been given to object modeling (Baltsavias, 2004). These are the results of advances in computing technology and also the need to deal with an increasing number and variety of sensors. Extraction of building features in urban areas, as one of the focuses of interest, requires very high spatial-resolution images. Aerial images with three visible bands have been conventionally used as the unique data source. Recently, more spectral channels are added and space-borne sensors also provide comparably high spatial-resolution images. Additionally, the correlation of a stereo-pair of images is the conventional approach to derive the elevation data, known as a Digital Surface Model (DSM). Height information is a supplemental source for feature extraction.

However, the level of automation in processing spectral images is very low (Brenner, 2005) due to diverse spectral information. Derivation of 3D information from a pair of spectral images is even more complicated. Suveg and Vosselman (2005) recommended that fusion with other data sources could reduce the complexity of extraction and then reconstruction. Alternatively, the development of Light Detection And Ranging (LiDAR) technology (Baltsavias, 1999; Wehr and Lohr, 1999) introduces a better option to improve the level of automation. A LiDAR surveying flight provides dense LiDAR point clouds that present well the complex structure of building features along with a very high-spatial resolution image. Such dense point clouds used as the primary data can successfully delineate the building footprints and structures. Solely using LiDAR data or its fusion with other data sources has been the dominant trend of research and practice in feature extraction (Haala and Brenner 1999; Maas and Vosselman 1999; Rottensteiner and C. Briese 2003; Vosselman et al. 2005).

To accommodate the complex scenes of urban areas, hierarchical and multi-scale approaches seem to offer promising solutions and recent research shows their more frequent use (Baltsavias, 2004). The work of Hofmann et al. (2002), Matikainen et al. (2001), and Vosselman et al. (2005) are typical examples of using a hierarchical bottom-up approach. However, it is hard to find a suitable set of parameters to determine the homogeneous regions. Scale-space processing might offer a solution to mitigate the difficulty. Investigating the behavior of objects across scale-space reveals important clues to guide the splitting and merging of homogeneous regions. The fact is that objects presented in an image possess scale properties. Exploitation of scale in image processing mimics human perception. Human perception ignores the details and groups the pixels into an object at a specific scale of observation.

The goal of this work is to construct a suitable scale-space processing framework for both extracting a building feature and storing it in a database. A specific (or generic) building model will not be used. Instead, by employing a multi-scale concept, a complex structure will be broken into several simple primitives across the scale-space and then easily re-linked in a multi-scale database. A brief description of area morphology (Vincent, 1992), which is employed to construct a nonlinear multi-scale scheme (Section 3), is presented in Section 2. The proposed method is demonstrated in Section 4 using data from a LiDAR surveying flight, followed by the conclusion and discussion for further developments in Section 5.

2 MORPHOLOGICAL SCALE-SPACE

Linear scale-space has been well developed in feature extraction and visualization (Lindeberg, 1993). To overcome the distortion problem in a coarser scale by the linear scale-space, which generates difficulty in linking across the scale-space, the nonlinear scale-space was proposed. The nonlinear scale-space keeps the main properties of a scale-

space like luminance conservation, geometry, or morphology (Petrovic, 2004). Generally, it performs a partition of an image into isolevel sets at each scale and links them with the closest one in the next scale. Our proposed method is based on the nonlinear scale-space employing area morphology (Vincent, 1992). The area morphology theory is as follows.

Let $x, y \in X$; x and y are *linked* for X if there is a path from x to y in X. X is *connected* if any x and y in X are *linked* for X. Y is a *connected component* of X if $Y \subseteq X$, Y is *connected*, and $Y = Z$ whenever $Y \subseteq Z \subseteq X$ and Z is connected.

The *connected opening* $C_x(X)$ of a set $X \subset M$, where $M \subseteq R^2$ at point $x \in M$ is the *connected component* of X containing x if $x \in X$ and $\varnothing$ otherwise.

Let $X \subset M$ and $s \geq 0$. The *area opening* of parameter s of X is given by

$$\gamma_s^a(X) = \left\{ x \in X \,\middle|\, Area\left(C_x(X)\right) \geq s \right\} \tag{1}$$

If $\{X_i\}$ denotes the connected component of X, it comes up with

$$\gamma_s^a(X) = \bigcup \{ X_i \mid i \in I, Area(X_i) \geq s \} \tag{2}$$

The binary area closing of parameter s ($s \geq 0$) of X is then defined as

$$\phi_s^a(X) = \left[\gamma_s^a(X^C) \right]^C, \tag{3}$$

where X^C denotes the complement of X in M.

Vincent (1992) then extended the definition of binary area opening or closing to grayscale area opening or closing. A grayscale image can be defined as a mapping $f : M \to R$.

The grayscale opening of f is given by

$$(\gamma_s^a(f))(x) = \vee \{ h \leq f(x) \mid x \in \gamma_s^a(T_h(f)) \}, \tag{4}$$

where $\vee$ stands for supremum, i.e. a lowest upper bound, and T_h is the threshold of f at value h:

$$T_h(f) = \{ x \in M \mid f(x) \geq h \}. \tag{5}$$

In other words, the thresholds of image M at all the possible values of h are firstly taken and the binary opening of each threshold set, $\gamma_s^a(T_h(f))$, is found. Subsequently, $\vee$ is applied to all the recently found $\gamma_s^a(T_h(f))$. It is similarly extended to grayscale closing by duality. Area morphological filtering does not depend on the shape of structural elements like classic morphological filtering. Therefore, it can effectively remove noise and simultaneously retain thin or elongated objects.

Applying area opening followed by area closing with a parameter s (AOC operator) on an image is like flattening this image by parameter s. This performance segments an image into the flat zones of similar intensity or isolevel sets, in other words. Therefore, iteratively applying an AOC operator with increasing s can generate a scale-space. A scale-space can be generated with an infinite number of scales. For the discrete dimension of an image, the number of scales increases by one each time a window (area) size increases from one pixel to the image size. However, it is time-consuming to be concerned with all the area values. Since objects in a scene tend to group into a limited number of size ranges, several size values may be enough for scale-space object extraction. Horizontal and vertical granulometry analyses (Vincent, 1994) can be used to find the potential sizes contained in an image. Alternatively, in regards to extraction of building features in a built-up area, there are several ranges of building sizes defined in the construction regulations. These ranges of sizes can be used as the parameters.

Let $\{s_i\}$, $i = 1, 2, \ldots$ be the parameter set to generate the scale space. The AOC scale-space $\{SP\}$ of an image M generated with $\{s_i\}$ is:

$$\{SP\}(M) = \left\{ SP_i \,\middle|\, SP_i = \phi_i(\gamma_i(SP_{i-1})) \right\}, \; i = 1, 2, \ldots, I \tag{6}$$

where $SP_0 = M$, ϕ_i and γ_i is closing and opening with s_i

SP_i, therefore, consists of the objects, e.g. building parts, tree, etc., greater than s_i. Fig. 1 illustrates three scales (s = 9, 300, and 1200) of the AOC performance on a grayscale image. The small details disappear in the coarse scales and an object can be formed in a specific scale depending on its size. For instance, a small house at the bottom-left of the image was fragmented into two objects in scale s = 9, it was totally formed when s = 300 and disappeared when s = 1200.

Fig.1.

Across the scale-space from a coarse scale to a fine scale, an object such as a building part follows the process of creation and split. Depending on the criteria based on several of its properties such as the height or spectral index, a newly created object might be a child of the current object that this new object falls into. Otherwise, the current scale is called the "*root*" scale of this object. An object can be extracted in its "*root*" scale. The links between objects across the scale-space are depicted in Fig. 2. Let us examine a three-scale-space such that *S1* is the coarsest one and *S3* is the finest one. In the current scale *S2*, there are two newly created objects named *A* and *B*. While *A* has similar properties to the bigger one in scale *S1*, *B* has different properties. As a result, *B* can be extracted in this scale *S2* with its two-level tree and *A* is associated with its father in *S1* and two children in *S3* to form a three-level tree. Those trees form a father-child relationship across the scale-space.

Fig. 2.

3 MULTI-SCALE IMAGE ANALYSIS

3.1 Pre-processing

This step transforms the elevation and spectral sources into a suitable format for the multi-scale analysis. The nDSM (normalized DSM), which is computed by taking the difference between the original DSM and the Digital Terrain Model (DTM), also called a bare-earth surface, is the first required input data. It represents the height of all overlying features such as buildings and trees on the 0 m flat terrain. An ortho-image which is often acquired simultaneously with the LiDAR point cloud on a surveying flight is the second input data. It is normally a 3 or 4-band image but there is no limit on the number of the multi-spectral bands. The required processing for the LiDAR point clouds and ortho-image is described below.

First, a LiDAR point cloud is interpolated into a grid format. The nearest neighbor interpolation method is employed here to preserve the sharp leap in elevation along the edges of buildings. The cell size is chosen as close as possible to the point density. The grid value now represents the elevation. Subsequently, we employed the wavelet-based clustering method (Vu and Tokunaga, 2004) to classify the LiDAR points into terrain and off-terrain points. Other classification methods can also be used here. The nDSM is then generated using the aforementioned method. Since we focus on building features, all the pixels lower than 2m in this nDSM are masked out. This is also to mitigate the possible errors caused by the LiDAR point classification. Instead of using the elevation value, there is a possibility to use the number of stories, which helps to simplify the building structure. The transformation from elevation value to the number of stories is simply implemented by an elevation slicing with the height of a story, about 2-3m, as the threshold.

Second, the spectral information is prepared. To utilize the pulse intensity as the near-infrared band, the pulse intensity is interpolated into a separate grid with the same cell size as that of the nDSM. Due to the noisy pulse intensity (Vosselman, 2002), the bilinear or cubic-spline interpolation method is recommended to reduce noise. The ortho-image is resampled into the nDSM spatial resolution, which is often coarser than the ortho-image resolution. The newly formed 4-band ortho-image will be used to spectrally

classify the objects. In feature extraction, the nDSM is used as primary data with the complement extracted from the spectral information. This is because the nDSM is able to better represent the building structures while the spectral information is diverse for the roofs of buildings.

3.2 Building extraction

The flowchart of the proposed scheme is depicted in Fig. 3. The processing is carried out separately for the height and spectral information prior to the final integration to develop the multi-scale database.

Fig. 3.

Regarding the height processing, a morphological scale-space is initially generated for the nDSM. As mentioned in Sec. 2, there are two possible ways to determine the scale parameter. It is noted that the finest scale and the coarsest scale are chosen corresponding to the smallest and the biggest buildings in the area, respectively. Subsequently, the father-child tree across the scale-space is constructed to prepare for the extraction. It is organized into two steps as described in the following paragraphs.

First, on each scale, every cluster defined as a group of pixels having the same value, e.g. the number of stories, are assigned their *ID* number. Additionally, other attributes are also assigned such as *H* (number of stories), *X0* and *Y0* (the starting point), *Xcen* and *Ycen* (the center point), *SCALE* (the scale at which it exists) and *AREA* (its size). The scanning across the scale space from a finer to the next coarser scale is subsequently carried out to eliminate the possible duplicate clusters since the chosen scale parameters might not be able to cover all the object sizes. If there are two overlapping clusters named *F* on the finer scale and *C* on the coarser one, *F* will be removed when *F* and *C* have the same number of stories. Then, if the values of those *F* and *C* are different, but *F*'s area is approximately equal to *C*'s area, *F* is also eliminated. The decision is made based on a defined threshold. For instance, it can be 0.9 or 0.85 which means *F*'s area is larger than 90% or 85% of *C*'s area. Second, the father-child tree is constructed across the scale-space. The link will be made not only between two consecutive scales but amongst all the scales since a large building might comprise a few very small pieces on its roof. Based on the falling of a small object *S* on a fine scale into a larger object *L* on a coarser scale, the *SUPID* of this object *S* is the *ID* of *L* and *SUPSCALE* is the scale on which *L* exists. Therefore, an object whose *SUPID* equals 0 is in the "*root*" scale.

Then a 4-band ortho-image is exploited to spectrally classify different types of objects. The main purpose is to remove the vegetation that might have a similar height as building features. Moreover, it is expected to further classify different roof materials. K-mean clustering is chosen for its simplicity and robustness. The clustering is applied to the areas of higher-than-one-story defined by the nDSM. This is to ensure that the parking or open spaces, the streets, etc., which have similar spectral signatures as the ones of the building roofs, are excluded. Prior to this clustering, the area morphological operator is employed as a filtering operator. In this way, the small meaningless objects are removed and the diversity of spectral information is also reduced. The above-implemented function for scale-space generation is used here with only one scale. This scale is the area that is a bit smaller than the smallest building. After clustering, the spectral indices of a defined number of classes are assigned. The corresponding information class such as vegetation, water, concrete, etc. to each spectral index is subsequently determined by the operator.

Spectral blobs are then used to adjust the extent of the above extracted building objects. They are used to remove the possible vegetation close to a building, which was mistakenly connected to the building object. To do that, the extracted objects from nDSM containing a vegetation blob are picked up for investigation. For each of those objects, a buffer of one pixel is generated along its edges by morphological dilation. The vegetation blobs which intersect that buffer are marked with 0 if they are completely contained in

the buffer or with 1 if they are not. All the intersected pixels between a building object and its vegetation blob of mark 1 are removed. As a result, the building object has been reshaped. It is to be noted that the vegetation blobs of mark 0 will be automatically removed shortly as only relevant spectral blobs to building features are retained in the following step.

Subsequently, all the spectral classes that are not related to building features such as vegetation and shadow are removed. The spectral index is assigned as the *SPE* (spectral information) attribute of the building as follows. On each scale, each extracted object from height data is crosschecked with the ones extracted from spectral data. Let $O_{i,k}$ be the currently checked object, i.e. object i in scale k. If there is no extracted object from the spectral data within the boundary of $O_{i,k}$, $O_{i,k}$ is removed from the database. Otherwise, the most frequently occurring of the spectral indices falling within the boundary of $O_{i,k}$ becomes the spectral attribute of $O_{i,k}$. This attribute is named *Speindex* in the building database. By setting a threshold about 40-60% in assessing the most frequently occurring spectral indices, the spectral blobs are then used to split a building if two different spectral classes are dominant in this part.

It should be noted that the rather low density LiDAR point cloud might be unable to effectively represent all the cases in a complex urban scene. Thus, in scale-space analysis, the adjacent extracted objects may share the same "fake" father on the coarser scale if the separation between them is too narrow. To mitigate the error due to this problem, the new index *StdDis* is introduced, which stands for standard deviation of distance. Since all the extracted objects have their center coordinates, it is easy to compute the distances from the centers of all the children to the center of their father. The standard deviation of the distances' distribution is the *StdDis* of the concerned "father". If its *StdDis* is larger than a certain threshold, implying that its children are dispersed in it, this father is a "fake" father. This step is to check all the objects in their "*root*" scale that have more than two children since they have a high possibility of being "fake", and is carried out prior to the conversion to vector format.

Finally, the extracted buildings are converted to vector format for input to a building database. Raster-to-vector conversion includes the detection of building edges and the arrangement of edge vertices to construct a polygon. Then all the polygons are stored in a shapefile (*.shp). The zigzag extracted boundaries from raster data are adjusted by a simplification algorithm like Douglas and Peuker's algorithm (Douglas and Peuker, 1973). All the listed attributes above are also attached in the conversion.

3.3 Multi-scale building database

The multi-scale extraction scheme, on the one hand, simplifies the extraction and construction task. On the other hand, it guides the development of a multi-scale building database. The components of each building are separately stored in their scales. It can be easily linked through the father-child relationship, i.e. through *SupID* and *Supscale* attributes. Thus, a very complex structure can be stored as well as matched with an existing database by extracting only the father polygons. This idea is demonstrated through an example in Fig. 4.

Fig. 4.

4 TESTS AND RESULTS

4.1 Data used

In order to demonstrate the proposed method, data from a LiDAR surveying flight captured over a densely built-up area of Roppongi, Tokyo, Japan, was used. The average LiDAR point spacing was about 1 m. The point clouds then were interpolated to a 1 m resolution surface model (DSM). Since the method was developed for all kinds of elevation data sources that can represent a surface model (DSM), it uses only the first echo of the LiDAR hits. The laser pulse intensity was merged with the ortho-photo to be

used as a near-infrared channel in segmentation. The selected test site covered an area of 900 x 650 m^2. A true color ortho-photo with spatial resolution of 20 cm was also provided as shown in Fig. 5. In this area, there were many high-rise and complex structures interspersed in the areas of tightly standing small and low houses. As a close-up shown in Fig. 5b, it is clear that it is extremely difficult to extract the building features solely from such an ortho-photo.

Fig. 5.

4.2 Results

The processing followed the general flowchart depicted in Fig. 3. All the processing steps were implemented by IDL coding (http://www.ittvis.com). The routines built include LiDAR point classification (Vu and Tokunaga, 2004), scale-space generation, cross-scale link, spectral classification, cross-scale reconstruction and raster-to-vector conversion. The nDSM prepared for multi-scale fusion is shown in Fig. 6c. It was computed from the original DSM (Fig. 6a) and the filtered DTM (Fig. 6b). Fig. 7 demonstrates the clustering of elevation blobs in the scale-space. The scale parameters shown in this figure were s = 50 (Fig. 7a), 200 (Fig. 7b), and 1000 (Fig. 7c). While the range from 50 to 200 pixels was suitable for most stand-alone houses and apartments, the range from 200 to 1000 pixels was suitable for commercial buildings. However, more intermediate s values like 100 and 500 pixels were also used to accommodate the building parts of various sizes. Also, 15 classes were called as the input parameter for spectral classification. The classified results showed that the spectral class ids from 4 to 15 were related to the building features.

Fig. 6.
Fig. 7.

The result of the entire area placed on the filter DTM, where brighter color means higher elevation, is shown in Fig. 8. The extracted building polygons are represented according to their height and scales. Brighter color shows smaller buildings. It should be noted that the processing here ignored the objects lower than 2 m. Thus, the open spaces, parking lots, roads, etc. were mostly excluded. However, elevated highways could not be excluded.

Fig. 8.

4.3 Accuracy assessment and discussion

The building footprints were manually digitized to use as the reference data. To be comparable, the manual digitization was carried out on the LiDAR-derived DSM and simultaneously crosschecked with a 1 m ortho-image. The ambiguity was solved by checking the existing GIS building database of Tokyo Metropolitan. Two very different areas were chosen for detailed comparison among the DSM, the reference data and the extracted results by the proposed method. The ortho-images of the selected areas are shown in Fig. 9. While the first area (Fig. 9a) is a complex area with diverse building types of various heights, the second one (Fig. 9b) is comprised of densely-built small houses with similar height. The difference is illustrated through the statistics of elevation points in Table 1.

Fig. 9.
Table 1

Both areas were completely extracted as shown in Fig. 10 which depicts the overlay of extracted building boundaries on the original DSM. Visually, the extracted results are reasonably well-matched with the reference and the DSM. Scale-space analysis allows

the simple but good visualization of the building structures. Since the reference data was manually digitized, it seemed to be more generalized and larger than the extracted results, and vice versa, the extracted results represent more detailed structures.

As shown by the reference data, there are 54 buildings in Area 1 and 60 buildings in Area 2. Visually checking, there was no commission error in both areas and no omission error in Area 1 whereas two very low houses were missed in Area 2. The main difference between the extracted results and their corresponding reference is the merging of two or more adjacent buildings, which have similar heights, into a bigger one. There are four cases in Area 1 and seven cases in Area 2. There is also a reverse case in Area 1 where the extracted results showed two buildings whereas the reference data claimed that they merged into one building. Object-based completeness and correctness were then measured to report the accuracy of extraction in a quantitative way as shown in Table 2. A set of criteria based on object orientation, shape and size was defined by an independent operator to judge whether an object were correctly extracted. The report confirmed that lower accuracy achieved in extraction of low and small houses in Area 2. It is expected that these errors could be reduced with a higher density LiDAR point cloud.

Table 2

Further quantitatively comparing in 2D, two shape similarity measures known as the relative arithmetic difference and the total relative shape dissimilarity (Henricsson and Balsavias, 1997) were employed as shown in Table 3. The average building sizes in Area 1 and Area 2 are approximately 200 m^2 and 76 m^2 respectively. Thus, the absolute area differences of 30.64 m^2 in Area 1 and 14.02 m^2 in Area 2 are both around 17% as represented by the relative arithmetic difference value. This value presents the overall difference in area and does not show under-segmentation or over-segmentation errors. These errors are represented through “A not B” (under-segmentation) and “B not A” (over-segmentation). In both test areas, the under-segmentation is worse than the over-segmentation. This confirms the earlier visual inspection proving that the reference data is more generalized. Total relative shape dissimilarity as a combination of under-segmentation and over-segmentation indicators, therefore, can represent the total shape difference in the areas. Both areas show similar values of shape dissimilarity. Using the computed area values, pixel-based completeness and correctness assessments reported good accuracies in both areas (Table 3).

Table 3

The four spectral bands used in classification were further checked to see whether there is some cue for roof material classification. To do this, the roof material of 167 buildings in the study area was collected by a field survey (Hasegawa 2005). The survey points were then matched to the extracted buildings by corresponding spatial location. The relationship between the spectral class ids and the roof material is illustrated in Fig. 14. It seems that the roof material “metal” that dominates classes 10, 12, 13 and 15 can be clearly identified. In addition, classes 7 and 9 might represent the roof material “cement”. The other two roof materials “tile” and “slate” show an unclear relationship. The noisy pulse intensity and the rather low-density LiDAR point clouds influenced this result.

Fig. 10.
Fig. 11.
Fig. 12.

The approximate 1 m point spacing of LiDAR point clouds used was rather coarse for the test area. In some dense areas with low houses, it was unable to separate those houses. The ortho-image was in fact provided at 20 cm spatial resolution, but to use the pulse intensity, we had to use 1 m spatial resolution. The noisy pulse intensity played an important role in removal of vegetation but was inapplicable in classification of the roof

material. A point spacing of less than 50 cm and more spectral bands are recommended for a better-extracted result. In addition, a generalization rule should be applied to the buildings in "*root*" scale, i.e. the footprints, for a better match with the existing 2D GIS database.

This study was initiated by the demand for faster development of building inventory techniques in Japan. Its goal, therefore, is different from the majority of research on the uses of LiDAR and image data, which focuses on 3D reconstruction for 3D city modeling and visualization. The scale-space solution has proved to be well-suited to the situation of many Japanese cities, especially for small objects on roofs as well as complex structures. This technique could therefore probably be successfully deployed in mapping other urban forms in, for example, European cities where each building stands at a distinguishable distance from the others.

The multi-scale approach developed here is most comparable to a processing scheme developed by Hofman et al. (2002). The major difference is that Hofman's approach relies on the eCognition software package, which perhaps limits room for improvement. Moreover, additional processing has to be developed to utilize the detected building outlines from eCoginition's result for either developing a building database or 3D building reconstruction. In contrast, the approach developed in this study is specifically designed for a 3D building database and the outcomes directly serve that ultimate goal.

The test with data acquired over Tokyo has demonstrated the deployment of the proposed method for building extraction which is an important step for map updating as well as for developing an inventory database for disaster damage detection. Since scale-space is generated based on the size of the objects, if an object of interest is distinguishable from the neighborhood then the approach would be applicable. Thus, if LiDAR data with a high density of points acquired over a forest area allows for the discrimination of separate tree crowns, this method could work well for applications in forest inventory.

5 CONCLUSION

A multi-scale scheme for building extraction was developed and demonstrated using the products of a LiDAR surveying flight. It is a fully automated approach making use of both spectral and elevation data. It can work well with any nDSM and spectral data sources. The multi-scale framework is not only the core for extraction but also for developing a building database. The multi-scale relational database is useful for storing a very complex structure and flexibly accessing matches with existing databases. Tests of this method using data captured over an area of Tokyo, Japan, gave good result in terms of completeness accuracy and the capability to accommodate diverse building types. However, higher density LiDAR data is recommended for use in such a dense area. More spectral information is required to run an automated derivation of the roof material. Since the processing runs on several scales, it seems logical that running time could be reduced using parallel processing. These two points will be explored in future work.

References

Baltsavias, E. P., 1999. Airborne laser scanning: existing systems and firms and other sources. ISPRS Journal of Photogrammetry and Remote Sensing 54, 164-198. doi:10.1016/S0924-2716(99)00016-7.

Baltsavias, E. P., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. ISPRS Journal of Photogrammetry and Remote Sensing 58**,** 129-151. doi:10.1016/j.isprsjprs.2003.09.002.

Brenner, C., 2005. Building reconstruction from images and laser scanning. International Journal of Applied Earth Observation and Geoinformation 6**,** 187-198. doi:10.1016/j.jag.2004.10.006.

Douglas, D., Peuker, T., 1973. Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. The Canadian Cartographer 10, 112-122.

Haala, N., Brenner, C., 1999. Extraction of buildings and trees in urban environments. ISPRS Journal of Photogrammetry and Remote Sensing 54, 130-137. doi:10.1016/S0924-2716(99)00010-6.

Henricsson, O., Baltsavias, E., 1997. 3-D building reconstruction with ARUBA: A qualitative and quantitative evaluation. In: Grun, A., Baltsavias, E., Henricsson, O. (Eds.), Automatic Extraction of Man-made Objects from Aerial and Space Images (II), Birkhauser Verlag, Basel, pp. 65-76.

Hofmann, A., Maas, H.G., Streilein, A., 2002. Knowledge-based building detection based on laser scanner data and topographic map information. In: International Archives of Photogrammetry and Remote Sensing 34 (part 3A), Graz, Austria, pp. 169–174.

Lindeberg, T., 1993. Discrete Derivative Approximations with Scale-Space Properties: A Basis for Low-Level Feature Extraction. Journal of Mathematical Imaging and Vision 3(4), 349-376. doi:10.1007/BF01664794.

Maas, H.G., Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. ISPRS Journal of Photogrammetry and Remote Sensing 54, 153-163. doi:10.1016/S0924-2716(99)00004-0.

Matikainen, L., Hyyppa, J., Hyyppa, H., 2003. Automatic detection of buildings from laser scanner data for map updating. In: International Archives of Photogrammetry and Remote Sensing 34 (part 3/W13), Dresden, Germany, pp. 218–224.

Petrovic, A., Divorra Escoda, O., Vandergheynst, P., 2004. Multiresolution segmentation of natural images: From linear to non-linear scale-space representations. IEEE Trans. on Image Processing 13 (8), 1104-1114. doi:10.1109/TIP.2004.828431.

Rottensteiner, F., Briese, C., 2003. Automatic generation of building models from LiDAR data and integration of aerial images. In: International Archives of Photogrammetry and Remote Sensing 34 (part 3/W13), Dresden, Germany, pp. 174-180.

Suveg, I. Vosselman, G., 2004. Reconstruction of 3D building models from aerial images and maps. ISPRS Journal of Photogrammetry and Remote Sensing 58, 202-224. doi:10.1016/j.isprsjprs.2003.09.006.

Vincent, L., 1992. Morphological Area Opening and Closings for Greyscale Images. In: Proceeding NATO Shape in Picture workshop, Driebergen, The Netherlands, pp. 197-208.

Vincent, L., 1994. Fast Grayscale Granulometry Algorithms. In: Proceeding EURASIP Workshop ISMM'94, Mathematical Morphology and its Applications to Image Processing, Fontainebleau, France, pp. 265-272.

Vosselman, G., 2002. On the estimation of planimetric offsets in laser altimetry data. In: International Archives of Photogrammetry and Remote Sensing 34 (part 3A), Graz, Austria, pp. 375-380.

Vosselman, G., Kessels, P., Gorte, B., 2005. The utilisation of airborne laser scanning for mapping. International Journal of Applied Earth Observation and Geoinformation 6**,** 177-186. doi:10.1016/j.jag.2004.10.005.

Vu, T.T., Tokunaga, M., 2004. Filtering airborne laser scanner data: A wavelet-based clustering method. Photogrammetric Engineering & Remote Sensing 70, 1267-1274.

Wehr, A., Lohr, U., 1999. Airborne laser scanning – an introduction and overview. ISPRS Journal of Photogrammetry and Remote Sensing 54, 68-82. doi:10.1016/S0924-2716(99)00011-8.

Table 1 Statistics of elevation points in the checked areas: the complex area (Area 1) and the area of densely-built small houses (Area 2)

	Area 1	Area 2
Mean (m)	38.32	32.89
Min (m)	22.11	24.01
Max (m)	77.67	53.06
Standard Deviation (m)	9.81	4.71

Table 2 Object-based completeness and correctness assessment of the complex area (Area 1) and the area of densely-built small houses (Area 2)

	Area 1		Area 2	
	Detected	*Reference*	*Detected*	*Reference*
Correctly extracted	39	43	31	44
Total objects	47	54	46	60
	Correctness = 82,98 %	Completeness = 79,63 %	Correctness = 67,39 %	Completeness = 73,33 %

Table 3 Comparison of the extracted result (B) and the reference data (A) by shape similarity measurement and pixel-based completeness and correctness.

Average value	Area 1	Area 2
Absolute difference: $\|A - B\|$ (m^2)	30.64	14.02
Difference between A and the intersection $A \cap B$: A not B (m^2)	32.49	17.42
Difference between B and the intersection $A \cap B$: B not A (m^2)	8.89	5.76
Relative arithmetic difference $\|A - B\|/A$ (%)	17.12	17.26
Total relative shape dissimilarity $(A \setminus B + B \setminus A)/A$ (%)	20.77	29.10
Pixel-based completeness $A \cap B / (A \cap B + A \setminus B)$ (%)	82.64	78.19
Pixel-based correctness $A \cap B / (A \cap B + B \setminus A)$ (%)	96.65	91.94

Fig. 1. Illustration of AOC performance with the change of parameter value *s*.
Fig. 2. Demonstration of the cross-scale link.
Fig. 3. General flowchart of the proposed multi-scale method.
Fig. 4. Demonstration of the multi-scale database structure.
Fig. 5. Ortho-image of the test area: (a) the whole area and (b) a close-up.
Fig. 6. (a) Original DSM provided by a LiDAR surveying flight. (b) the filtered DTM and (c) the derived nDSM of the test area.
Fig. 7. Illustration of object forming based on elevation across the scale-space from finer (a) to coarser scales (b) & (c).
Fig. 8. Perspective view of the extracted result for the entire area.
Fig. 9. Extracted ortho-images of two selected areas: (a) the complex area (Area 1) and (b) the area of densely-built small houses (Area 2).
Fig. 10. Extracted results overlaid on the original DSM: (a) Area 1, (b) Area 2.
Fig. 11. Extracted results overlaid on reference data: (a) Area 1 (b) Area 2.
Fig. 12. The relationship between the roof material and the spectral classes; metal, slate cement, and tile.

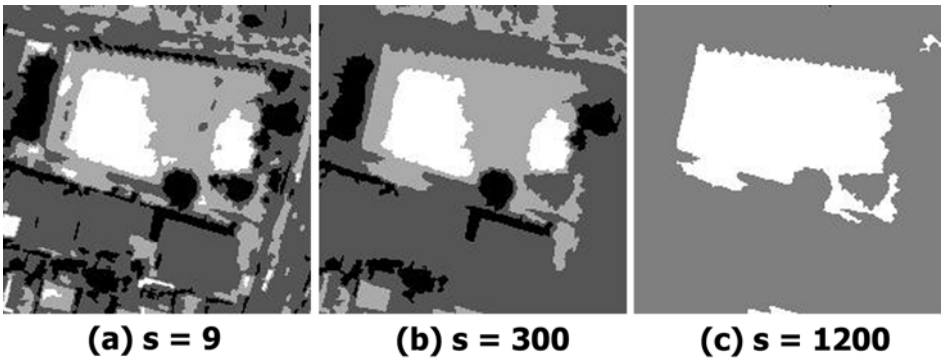

Fig.1

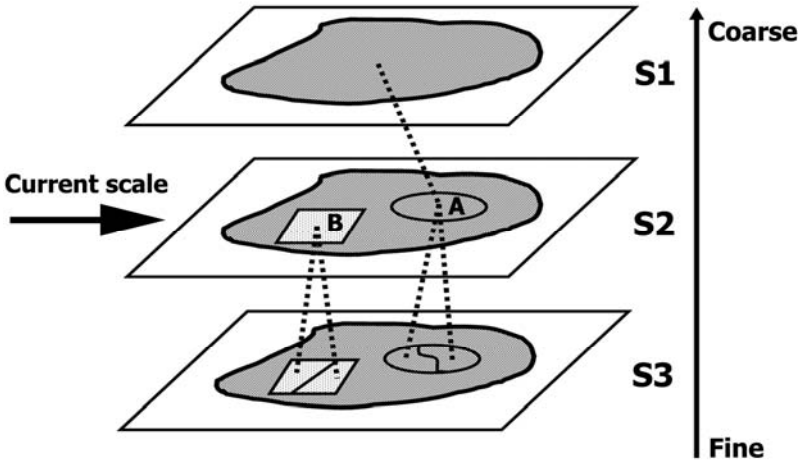

Fig. 2

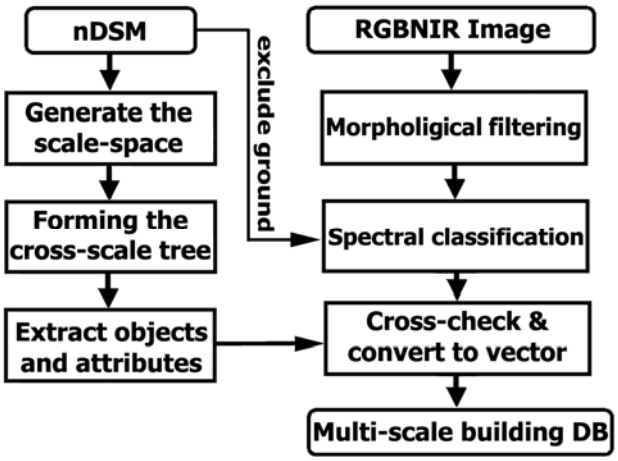

Fig. 3

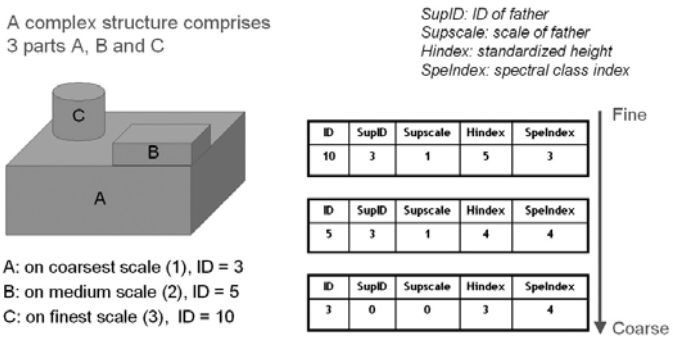

Fig. 4

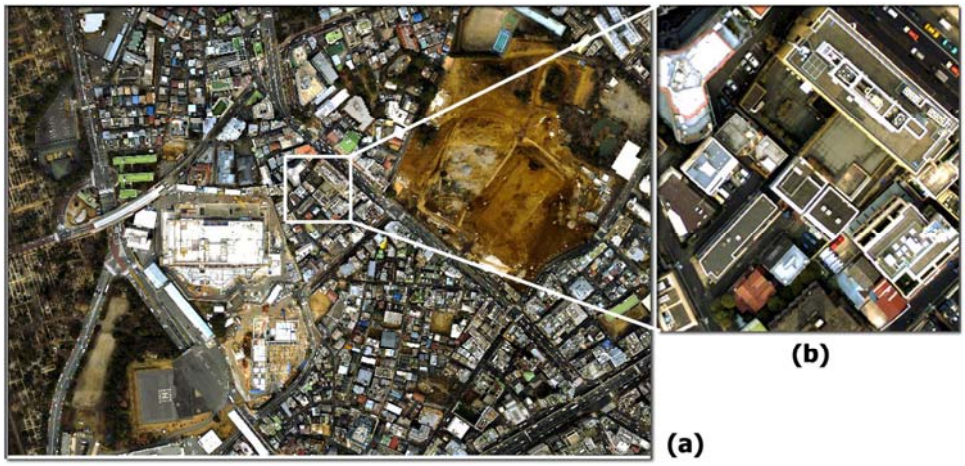

Fig. 5

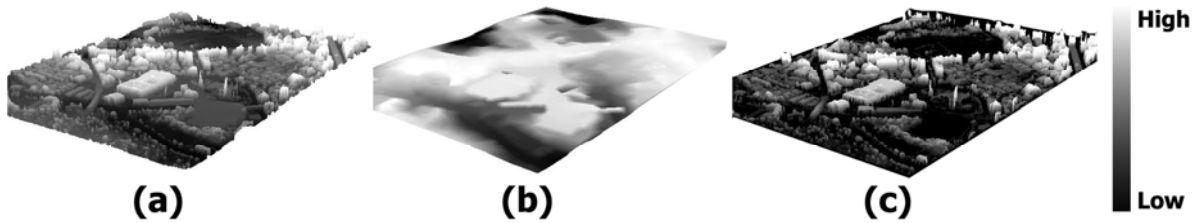

Fig. 6

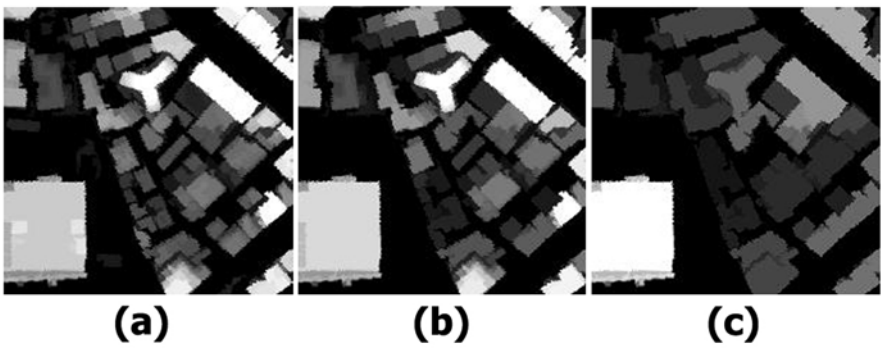

Fig. 7

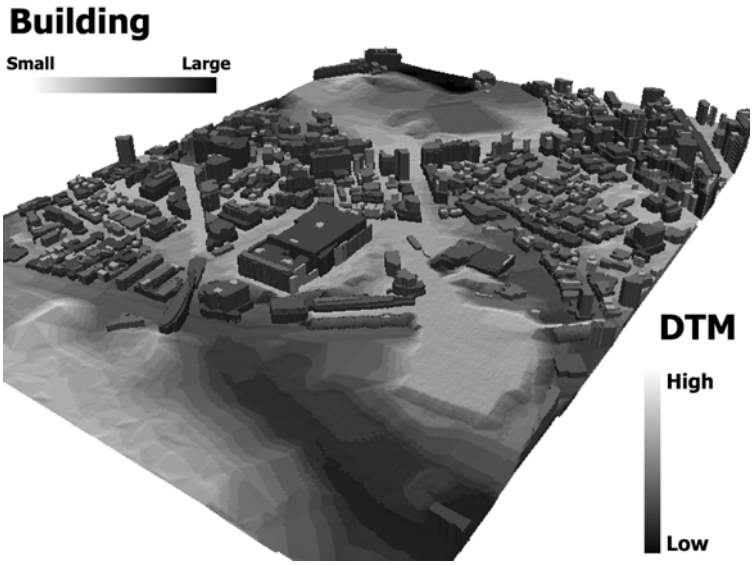

Fig. 8

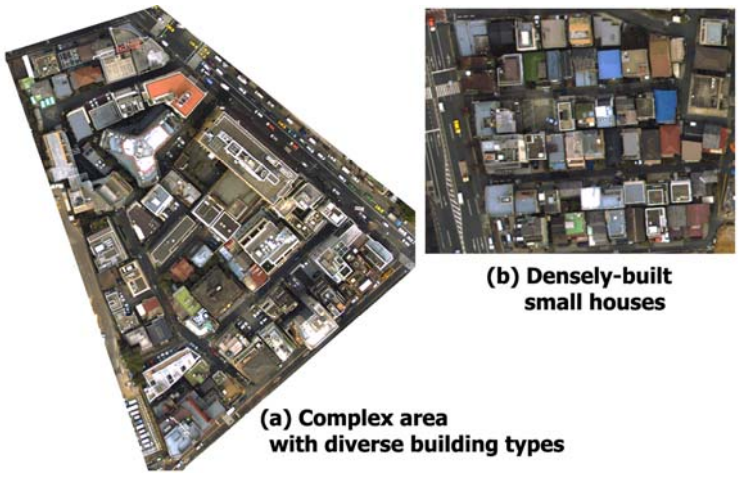

Fig. 9

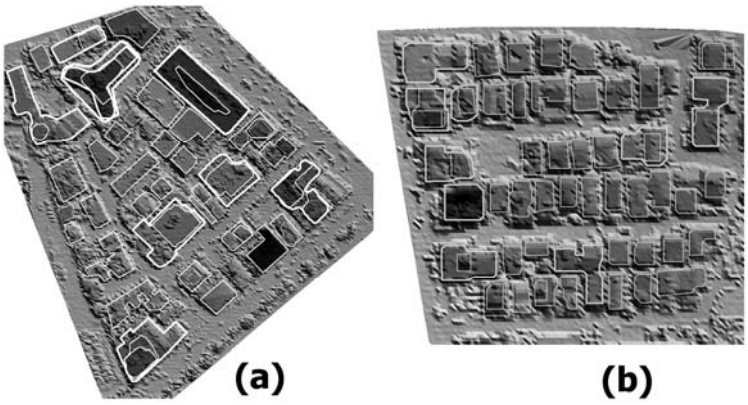

Fig. 10

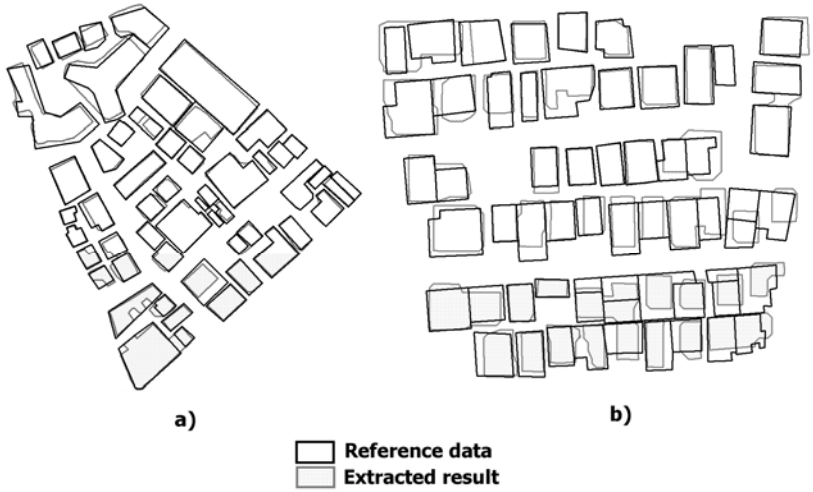
a)
b)
Reference data
Extracted result

Fig. 11

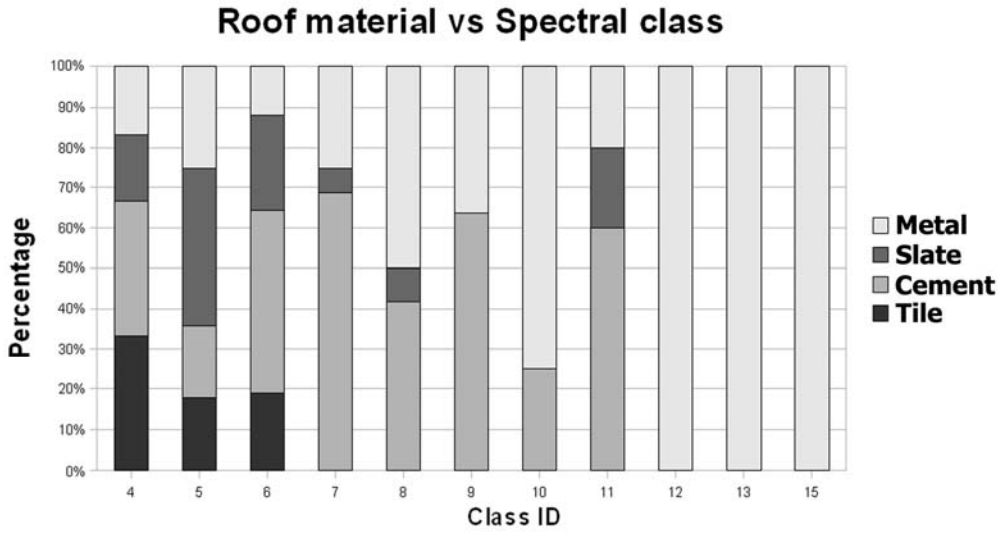
Roof material vs Spectral class
Percentage
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
4
5
6
7
8
9
10
11
12
13
15
Class ID
Metal
Slate
Cement
Tile

Fig. 12