VARIOUS BUILDING DETECTION METHODS WITH THE USE OF IMAGE AND LIDAR DATA

Nusret Demir

Original scientific paper

In this work, an automated approach for building detection using airborne images and LIDAR data is presented. A combined approach of four methods achieved the best results, using slope-based DSM filtering as well as classification of multispectral images, elevation data and vertical LIDAR point density. The first variant of building detection is based on multispectral classification and DSM filtering. In the second variant, DSM blobs, mainly consisting of buildings and trees, are detected by subtraction of the DTM from the DSM. The third variant uses the planimetric density of raw LIDAR DTM data to detect the above-ground objects. The fourth variant is like the third one, but uses the vertical density of the raw LIDAR data (all points) to distinguish trees and buildings. In the evaluation, the combination of the four methods yields 94 % correct detection at an omission error of 12 %.

Keywords: buildings, detection, DSM/DTM, LIDAR data, image classification, image processing, object extraction

Različite metode otkrivanja građevina pomoću snimaka i LIDAR podataka

Izvorni znanstveni članak

U ovom radu predstavljen je automatski pristup otkrivanja građevina uprabom zračne snimke i LIDAR podataka. Kombinirani pristup četiri metode postigao je najbolje rezultate, rabeći filtriranje DSM-a preko nagiba, i klasifikaciju multispektralnih snimaka, visinskih podataka i okomite gustoće LIDAR točaka. Prva varijanta otkrivanja građevina je utemeljena na multispektralnoj klasifikaciji i filtriranju DSM-a. U drugoj varijanti, objekti na DSM-u, a to su ponajprije građevine i drveće, otkriveni su pomoću računanja razlike između DTM-a i DSM-a. Treća varijanta rabi planimetrijsku gustoću neobrađenih LIDAR DTM podataka kako bi otkrila objekte iznad površine. Četvrta varijanta temelji se na trećoj varijanti, međutim rabi okomitu gustoću neobrađenih LIDAR podataka (sve točke) kako bi razlikovala drveće i građevine. Testiranje i ocjenjivanje postupka pokazalo je da kombinacija ove četiri metode vrši ispravno otkrivanje građevina u 94 % slučajeva uz grešku izostavljanja od 12 %.

Ključne riječi: DSM/DTM, izvlačenje objekta, klasifikacija slike, LIDAR podaci, obrada slike, otkrivanje, zgrade

1 Introduction

During recent years, building detection is addressed as one of the major problems in computer vision and photogrammetry. Buildings are important objects for many applications. Municipal planning, decision support systems, telecommunications, energy potential and efficiency studies, network planning, military applications, aviation purposes, flight simulations, any kind of navigation and mapping applications are all key sectors that require building objects. Telecommunication applications, as an example, require building models to assess and manage the wireless networks. Municipalities and cadastre offices need building models for tax and law purposes. Additional applications can be sentenced for similar case studies such as navigation, visualization, show-business, etc.

Accurate enough, complete or up-to-date building information is not always available. Thus, methods are needed for generation and updating of accurate and complete building data with a high degree of automation.

As input data, we use both digital images and LIDAR data. The proposed methods, through simplifications and slight modifications, could be employed using only image or LIDAR data. However, we believe that these data are partly complimentary and should be used in combination. E.g. digital images offer better image quality and spectral information and better edge (often surface discontinuity) definition, while LIDAR points have less and smaller gross errors, can provide a DTM everywhere (including the trees), provide multiple echoes per pulse and thus can also model the vertical structure of objects, e.g. trees. In the proposed methods, both DSMs and DTMs are used. These can come from LIDAR data or DSMs can be generated by image matching and then using DSM

filtering, as for classification of LIDAR points, a DTM can be derived. However, such a DTM will be inferior to a LIDAR DTM in case of trees, especially when covering a large area.

In the past, some research dealt with building detection by using only image data [1, 2, 3], and some others by using LIDAR data [4, 5, 6, 7]. The methods, which use single images, assume that the building objects have simple shape, so the results are very limited. Occlusions and shadows, although a general problem, also cause increased errors, when using one image. In general, the developed methods extract only simple buildings like flat-roof and gable-roof buildings. Secondly, a user interaction is generally required to derive accurate results. For the methods that use stereo or multi-images, generally, there are several parameters, which have to be tuned for the processing, or some manual interactions are needed to get robust results. The results are relatively accurate, especially for simple buildings. Since surface discontinuities in an image-based DSM are not so sharp with older matching methods, complex-shaped buildings are difficult to extract, so usage of additional LIDAR data can lead to better results.

In LIDAR-based methods, the results are mainly depending on the density of points. The cost of LIDAR data is generally lower than image-based DSMs, while the overall processing is faster. The capability of LIDAR data for detection of surface discontinuities is weaker compared to edges extracted from images [19]. In general, it is difficult to detect accurate building models with low-density LIDAR data. LIDAR and photogrammetry are considered as complementary to each other by [8]. The integration of both technologies is believed to lead to more accurate and complete products [8]. For example, image data have advantages by providing spectral

information, and LIDAR data have strengths to extract features using geometrical information and provide information on the vertical structure of semi-transparent objects like trees. There are several methods which use both LIDAR and image data ([9, 10, 11, 20]), and the research on combination of both datasets is still a hot topic, while existing approaches have some limitations in the full exploitation and combination of these datasets and their partial results. This study is an extended version of the detection work presented in [12].

1.1 Input Data

The test site is in Vaihingen, Germany. The dataset has been provided from the DGPF camera evaluation project ([13]). DMC digital images with NIR, R, G, B channels and raw LIDAR point cloud are used. Reference vector data has been collected by stereo measurement using the DMC images, with 10 cm accuracy.

1.1.1 Image Data

DMC images have 8 cm GSD. They have been acquired in August 2008

1.1.2 LIDAR Data

LIDAR points have a density of 5 pts/m^2. There is no explicit verification of the accuracy of the LIDAR data, it has been assumed that the vertical positional error (sigma) as 20 cm with vertical differences (18 cm) caused by residual errors in the individual system components: GNSS, laser scanner and the mounting [13]. The data have been acquired in July 2008.

2 Building detection

Fig. 1 shows the workflow for the building detection. The initial rationale for this procedure is to test and compare various simple approaches and then compare these results to their combination, which would be hopefully better.
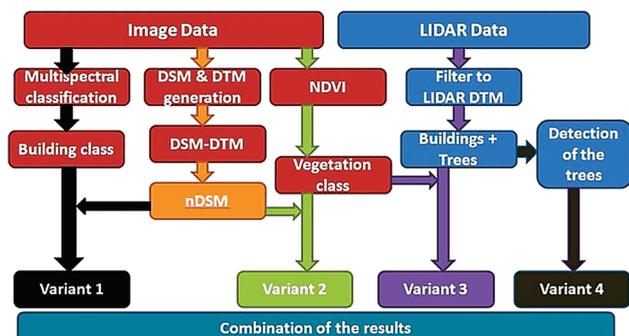
Figure 1 Workflow of the building detection, red indicates the items related to the image data, blue LIDAR data, orange is nDSM, and the results of different approaches have different colours

The buildings are detected by several variants, while each variant has different characteristics. A combination is proposed to achieve the best detection result. The approach starts with detection of above-ground objects in a DSM.

Then, the first variant uses the spectral information to detect the buildings; a pixel-based supervised classification is used to detect buildings. Using the height information from the nDSM, the mixed building and ground class could be separated.

The second variant does not use all spectral information but only NDVI and eliminates the trees from the detected above-ground objects. In this approach, ISODATA clustering of the NDVI image extracts the vegetation and no vegetation regions. ISODATA is an unsupervised classification technique, which segments the image into the number of input classes. Finally, the intersection of no vegetation regions and nDSM blobs extracts the buildings. Since urban areas have diverse spectral information, the buildings are detected in these two variants with using more image channels in multispectral classification, and red and infrared channels with using only NDVI. In the third variant, the raw LIDAR DTM points are used to detect the above-ground objects using the density of LIDAR-DTM data. Using the vegetation class from the NDVI classification as a mask removes trees to detect the buildings.

The fourth variant is fully based on the LIDAR data, so it does not include any drawbacks from image data (e.g. vegetation on roofs), but has weaknesses, especially on building outlines, which are often detected as trees.

2.1 DSM/DTM

Although LIDAR data is available in our test, we initially preferred to use an image-based DSM, which can always be generated densely with high-resolution images For the DSM generation, the NIR channel is chosen because of its better contrast on edge features. The noise in the images is reduced using an edge-preserving filter. Then, a Wallis filter is applied to enhance the contrast. A DSM is generated with 25 cm grid spacing using the matching method by [14]. A DTM is computed for extracting the above-ground objects. A slope-based progressive morphological filtering method ([15]) has been used to reduce the DSM to DTM. An opening (erosion + dilation) is performed on the DSM to gain a secondary surface. The elevation difference of a grid between the previous and current surface is compared to a threshold to decide, if a grid point is a ground measurement. The height difference threshold has been computed using the predefined maximum terrain slope as defined by Zhang [15]:

$$\mathrm{d}h_{T,k} = \begin{cases} \mathrm{d}h_{\max} & \text{if } \mathrm{d}h_{T,k} > \mathrm{d}h_{\max} \\ s\,(w_k - w_{k-1})c + \mathrm{d}h_0 & \text{else if } w_k > 3 \\ \mathrm{d}h_0 & \text{else if } w_k \leq 3 \end{cases},$$

where $\mathrm{d}h_{T,k}$ is the height difference threshold, $\mathrm{d}h_0$ is the initial elevation difference threshold which approximates the error of DSM measurements, $\mathrm{d}h_{\max}$ is the maximum elevation difference threshold (m), c is the grid size (m) and w_k is the filtering window size (in number of cells) at kth iteration.

2.2 Variant 1: Multispectral classification

Maximum likelihood classification (MLC) is one of the most widely used techniques in supervised classification [16]. In case additional information is available, e.g. DSM, DTM or ground plan, the result from the supervised classification can be improved using rules for merging similar classes or dividing mixed ones. A rule-based supervised classification is performed to determine the building class using ERDAS software package. Prior to performing the classification, the input classes are defined. Since the focused objects are buildings in urban area, there is no variety of input classes. In context of PEGASE project, the suggested input classes are the buildings, bare ground, road, trees, grass and shadow. After performing the classification with selection of ca. 4000 pixels for each class, the results are refined by splitting and merging the mixed classes to detect the final building objects. After generating the orthoimage, extra images are created from original spectral channels. These additional images are NDVI, saturation, and PCs.

NDVI is advantageous for detecting vegetated regions; PCs may contain additional information on the surfaces, which cannot be extracted using the original input channels. The saturation image is useful for identifying shadow regions much more accurately.

NDVI and saturation images are generated by input multispectral channels and RGB to IHS transform respectively. After including these derived images, the class separability is analysed using transformed divergence separability analysis for selecting the input channels among all available channels. The selected channels are NIR, R, first PC, NDVI, saturation. A maximum likelihood classification is performed, and the building class is defined with rules using nDSM and size criteria for the objects. This approach is the only one that is not fully automated, because of the training stage.

2.2.1 Refinement of the classification

The MLC method calculates the probability of a pixel and assigns it to the class with the highest probability. The disadvantage of this method is that it classifies each pixel regardless of how far it is from the mean value of the class. The results are therefore not always consistent; additional information sources need to be used and specific rules designed to refine the classification result. Some classes may have low separability (as shown in Fig. 2, the grass and trees are mixed, bare ground and buildings are mixed). Use of detected blobs is one option to separate such classes or to merge similar classes. As this study focuses on buildings, size can also be an additional rule to refine the results. The rules developed during experiments using a test dataset are:

- The classes which are not separated and at different heights can be split using the nDSM.
- The classes which are not separated and at the same height are merged using the nDSM.
- Some errors can be eliminated by using a size threshold. This threshold eliminates the detected objects smaller than 25 m^2.
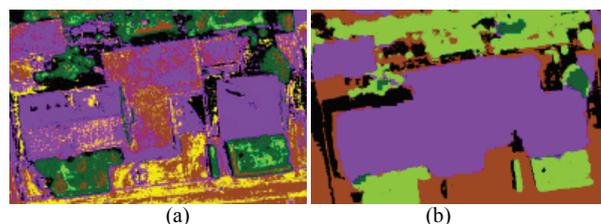
Figure 2 Refinement of the classification result: (a) Classification result, (b) Refined result: purple - building class, yellow - road, brown - bare ground, light green - grass, dark green - tree, black - shadow

Fig. 2 shows an example of the result of multispectral classification using the mixed classes. In the image on the right, the result is improved based on the rules for the refinement. Because of low separability between roads and bare ground, and between buildings and bare ground, these classes are not well separated and need to be refined.

Road and bare ground classes are merged, and bare ground and building classes are split using the nDSM. Road class is not an interest for the study, so merging it with bare ground class does not affect the results of building detection. As shown in Fig. 2, the class with yellow colour indicates road class and then it is merged with ground class (brown) if its height is below 1,5 m.

Finally, the ground class is changed to building class when its height is more than 1,5 m. As an alternative, the DSM (or nDSM) can be used as a channel in the classification, instead of using the nDSM in the refinement process. With this, the classes which are not separable will be able to be split using height, e.g. bare ground and buildings have different heights, so DSM has a potential to allow a better separability among the input classes.

2.3 Variant 2: Using nDSM and NDVI Classification

In this variant, NDVI and nDSM are used to obtain buildings. While the first approach requires a training stage, the second approach is developed to avoid any manual interaction. The vegetation class is extracted by ISODATA clustering [17] of the NDVI image. When investigating vegetated areas two classes are used, namely vegetation and non-vegetation.

After obtaining the vegetation class, the tree objects are identified through the intersection of nDSM and the vegetation class from the ISODATA classification of the NDVI image. By subtracting trees from the nDSMs blobs, mainly buildings are detected. Some errors are eliminated using a size threshold (25 m^2).

2.4 Variant 3: Using LIDAR DTM and NDVI

This approach uses both image and LIDAR data. Thereby, the above-ground objects are detected through LIDAR data and trees are eliminated using NDVI. Following elimination of the trees, the buildings are not the only objects (e.g. moving objects) that remain and, hence, size criteria are used to determine buildings. The main objective is the detection of above-ground objects in the LIDAR DTM data and the subsequent elimination of vegetated regions using the classification result from NDVI. The process starts with the identification of

regions, which have low density in the raw LIDAR DTM points.

The used raw LIDAR DTM points are data that include only those points that belong to the ground surface. This dataset has to be raw, and not interpolated. These kinds of data-sets can be directly acquired from mapping agencies or can be derived from DSM by using filtering techniques. In this variant, an external software SCOP++ [18] is used to obtain LIDAR DTM-AV in case no LIDAR DTM dataset is available e.g. for Vaihingen dataset. SCOP++ takes the LIDAR point cloud and delivers both LIDAR DTM point data with holes and an interpolated one.

The main indication for above-ground objects in the LIDAR DTM dataset is the density of the point cloud. As this dataset includes only points that belong to the ground surface, the areas of buildings are represented as voids and the density of vegetated areas is relatively low in comparison to the bare ground surfaces as illustrated in Fig. 3.
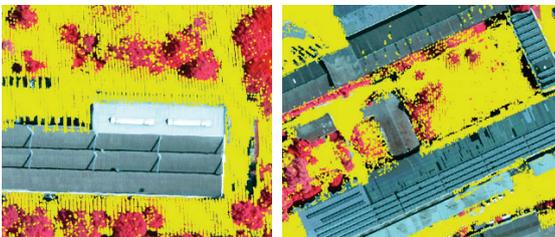
Figure 3 LIDAR DTM overlaid on an orthoimage

The local density is calculated to identify those areas that have above-ground objects. To calculate local density, a regular grid is first overlaid on the LIDAR DTM dataset. The cells are overlapped with half of the grid size. The cell size is chosen as 1 m for calculation of the density per square meter. Then, LIDAR points are counted in each grid cell. The total number of points in a cell gives us the local density. The algorithm is as follows:

- Overlay $W \times W$ cells (Gn) on LIDAR DTM
- Count the points in cell Gi
- If the total number of points is less than threshold T, Gi represents the areas which have low density.
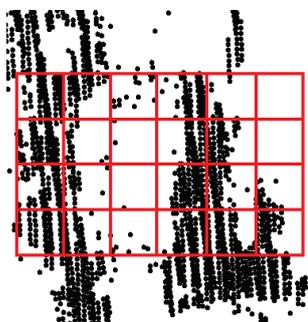
Figure 4 Density calculation over LIDAR DTM. Overlaid grids (red) and LIDAR DTM points (black).

The critical part is the selection of the threshold for the low-density estimation. The low-density threshold can be determined to represent half of the average density of the LIDAR dataset. After choosing the threshold, the areas that correspond to the low-density values are

extracted from the LIDAR DSM as above-ground objects. These extracted areas primarily represent trees and buildings.

The trees are eliminated with the vegetation class, which was extracted from a NDVI ISODATA classification. Some errors are eliminated using a size filter (25 m^2).

2.5 Variant 4: Detection of the Buildings from LIDAR Data

The detection of trees is not always successful when using image data, since the roofs may be covered by tree branches or they may contain a grass surface. Therefore, these surfaces are detected as vegetation using any image classification process. The use of LIDAR data avoids this problem since the detection is based on the geometry of the objects.

In this approach, the above-ground objects are detected with a density analysis of the LIDAR DTM points and extracted from the LIDAR data as described in variant 3. To detect the buildings, the trees are first detected and later removed from the previously detected above-ground objects.

Some assumptions are made to detect the tree objects in the LIDAR dataset:
1) Roughness of the surface: It is assumed that the points, which correspond to trees, cannot be fitted well with a planar surface, while the points on the building roof can.
2) Density in the vertical profile: The trees contain many LIDAR returns, which results in a high vertical density in the vertical profile over the tree objects. Building objects have points on the roofs and a few points on the walls. There are no points below the roof planes in the dataset (Fig. 5).
3) The planimetric density of LIDAR points at trees is higher than the one at the buildings. Additionally, the estimated minimum tree height (3 m) is taken into account.
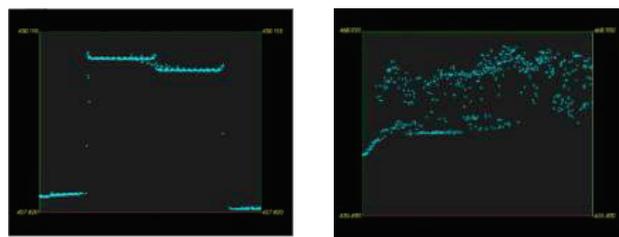
Figure 5 Profile view of the building (left) and tree object (right)
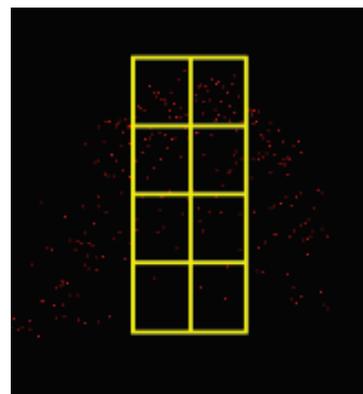
Figure 6 Sub-regions in vertical profile. Red: LIDAR points

The process is started with the detection of planar features in the LIDAR point dataset. The plane detection is employed using the RANSAC method (Schnabel et al., 2007). The points that fit a planar surface are identified and labelled. Then, a grid is overlaid on the dataset.

The grid size is chosen with consideration of the average occupied area by a single tree as 25 m². Secondly, the grid cells are subdivided into eight equal areas in height (Fig. 6) in the *xz* and *yz* planes.

For each grid

- The number of sub-regions (*SB*) that contain a LIDAR point is counted to define the vertical density of the search window.
- The total number of points (*NP*) is counted.
- The maximum height difference (*H*) between the points is calculated.

Then, the search window is classified as a tree, if all the following conditions apply for one of the *xz* or *xz* planes:

$$SB > TV \text{ and } NP > TD \text{ and } H > TH,$$

where *TV* (4 sub-regions), *TD* (15 points), *TH* (3 m) are thresholds for vertical density, horizontal density and minimum tree height, respectively. The selected numbers are found in parentheses. Following the detection of trees using these criteria, the trees are eliminated from the above-ground objects to derive the building objects.

3 Results

In first variant, the correctness and completeness measures are calculated as 94 % and 85 % respectively for 218 buildings in the reference data.
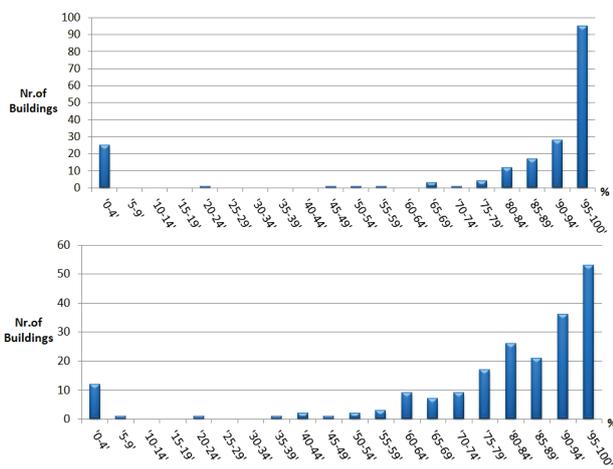
Figure 7 The histograms of the correctness (top) and the completeness (bottom) of the variant 3

Table 1 Quality values of the results from the variant 1

Correctness (%)	94
Completeness (%)	85
Number of detected objects in the building class	189
Detected objects which intersect with buildings in the reference data	163
Number of buildings in the reference data	218
Number of buildings in the reference data, which don't intersect with any detected object	10

The histogram for the correctness and completeness can be found in Fig. 7. The detected buildings are shown in Fig. 8.

Figure 8 Detected buildings (yellow) from variant 1

The quality values that include correctness, completeness, and information on the number of buildings, which are detected and missing, are shown in Tab. 1.

In the generated result, some of the buildings, mostly small ones, could not be detected because of the errors in filtering of the DSM. In the blob detection process, subtraction of the DTM from the DSM cannot produce the blob for these unfiltered surfaces. This is the main reason for the peak "0 ÷ 4 %" of the completeness histogram.

In second variant, the building detection result is shown in Fig. 9. The histograms for correctness and completeness are shown in Fig. 10. Tab. 2 shows the quality values for the results.

Figure 9 Detected buildings (yellow) from variant 2
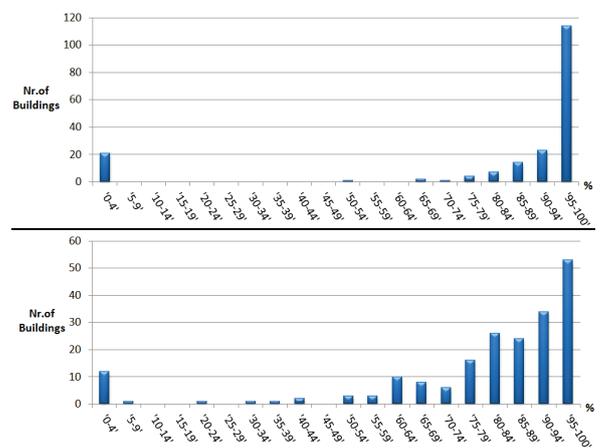
Figure 10 The histograms of the correctness (top) and the completeness (bottom) of the variant 2

Table 2 Quality values of the results from the variant 2

Correctness (%)	91
Completeness (%)	85
Number of objects in the building class	187
Detected objects which intersect with buildings in the reference data	165
Number of buildings in the reference data	218
Number of buildings in the reference data, which don't intersect with any detected object	9

The factors that cause errors are similar as in the first approach. However, in this approach, the shadowed regions (which are not treated as a separate class) in the vegetated regions are also detected as buildings, and that causes a lower correctness compared to the result of the first approach. This is the main reason of the peak "0 ÷ 4 %" of the correctness histogram.

In third variant, the detected buildings are shown in Fig. 11.

The correctness and completeness histograms are shown in Fig. 12, respectively.
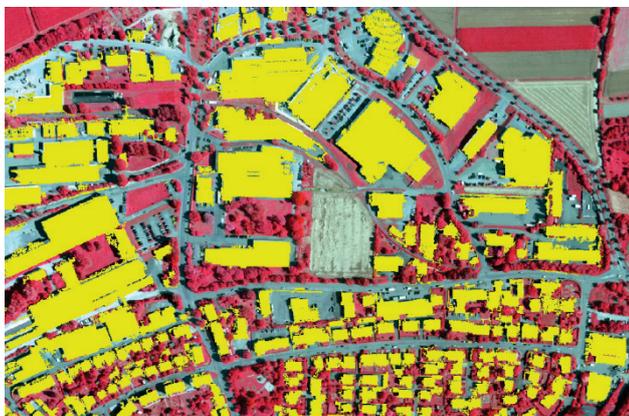
Figure 11 Detected buildings (yellow) from variant 3
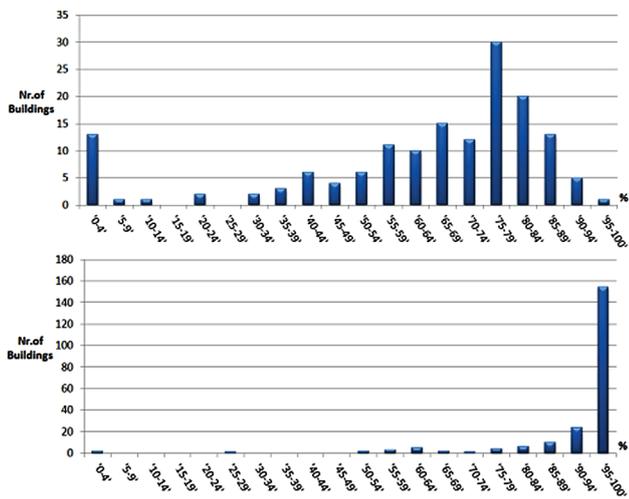
Figure 12 The histograms of the correctness (top) and the completeness (bottom) of the variant 3

The quality values are shown in Tab. 3.

The error sources are similar to a certain extent to the ones from the second variant, since they both use NDVI to eliminate the trees. Therefore, shadows on the trees produce false detections also in this approach. The high peak in the first column of the correctness histogram comes from these false detections.

However, as shown in the completeness histogram, only two buildings do not intersect with any detected building in the detection, as the LIDAR-DTM data is generated using a much robust method of SCOP++.

Table 3 Quality values of the results from variant 3

Correctness (%)	81
Completeness (%)	88
Number of objects in the building class	127
Detected objects which intersect with buildings in the reference data	114
Number of buildings in the reference data	218
Number of buildings in the reference data, which don't intersect with any detected object	2

In fourth variant, detected buildings are shown in Fig. 13 and correctness and completeness histograms are shown in Fig. 14, respectively.

Figure 13 Detected buildings (yellow) from variant 3
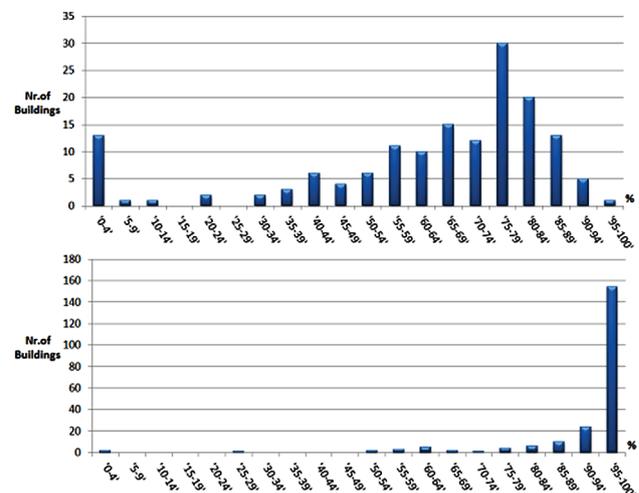
Figure 14 The histograms of the correctness (top) and the completeness (bottom) of the variant 4

The quality values are shown in Tab. 4.

Table 4 Quality values of the results from variant 4

Correctness (%)	80
Completeness (%)	80
Number of objects in the building class	267
Detected objects which intersect with buildings in the reference data	167
Number of buildings in the reference data	218
Number of buildings in the reference data, which don't intersect with any detected object	6

The correctness and completeness measures are calculated as 80 % for both measures. The trees which could not be eliminated from the above-ground objects, are detected as buildings and produce a high peak at "0-4" range of the histogram. Completeness is also low compared to the result from variant 3, because most walls are detected as trees and are eliminated from the building class.

3.1 Combination of the results

Binary operations are used for the combination of the results to increase the quality of the detection.

The combination of the results is performed based on consideration of the properties of the used datasets and the advantages and disadvantages of each approach. The intersection of the results will lead to higher correctness but less completeness, on the other hand, union of the results gives higher completeness but less correctness. Therefore, an optimal combination has to be investigated.

The results from the image data depend on the spectral information. Shadows and vegetation on the roofs may cause omission errors. In case only NDVI is used as in second variant of detection, there is a risk to detect the shadows on vegetation as buildings; so this produces high commission error. It is obvious that intersection of the results from the variants, which use image data, minimizes the errors.
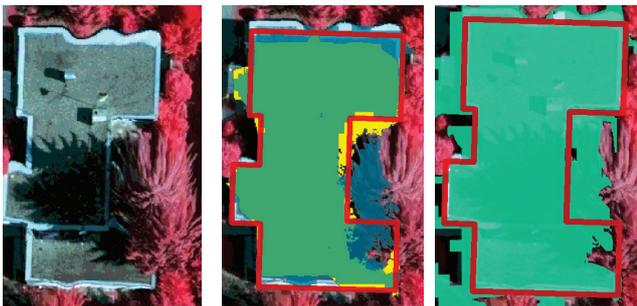
Figure 15 Combination of the results, Left: example on the orthoimage, middle: intersection of the variants 1 and 2. Red outline: reference data, blue: result from variant 1 yellow: result from variant 2 green: intersection of the result from variants 1 and 2, right: green: union of the results from variant 3 and 4, red outline: Reference data.

As shown in Fig. 15, the red colour indicates the result from the first variant, and the yellow one the result from the second variant. The intersection generally refines the border of the detected building, but it decreases the detected area. Furthermore, LIDAR-based results (variants 3 and 4) are used in the union process since they are not affected by the spectral problems mentioned above. With union of the result of variant 3 and 4, the detection region is enlarged and the completeness is increased. It is preferred to keep all buildings as much as possible with a small omission error. Therefore, a union process is applied between the results from intersection of 1 and 2, and union of 3 and 4.

Moreover, the final detection result is supposed to be part of the blobs. Thus, finally, an intersection is performed with the detected blobs. Therefore, the suggested binary operation is (R1 ∩ R2 ∪ (R3 ∪ R4)) ∩ Blobs.

Detected buildings are shown in Fig. 17 and correctness and completeness histograms are shown in Fig. 16, respectively.
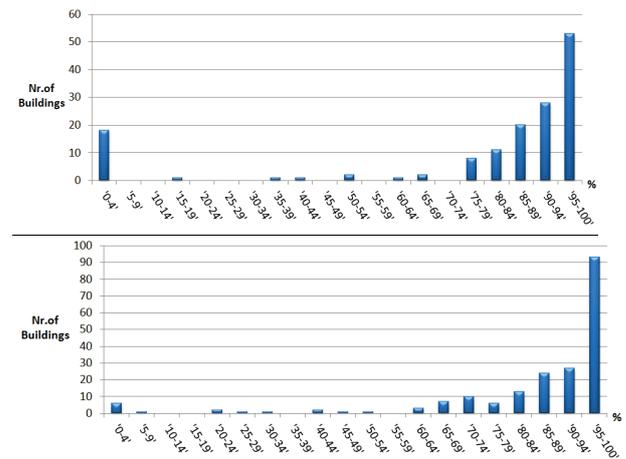
Figure 16 The histograms of the correctness (top) and the completeness (bottom) of the variant 4

Figure 17 Detected buildings (yellow) from combination of results
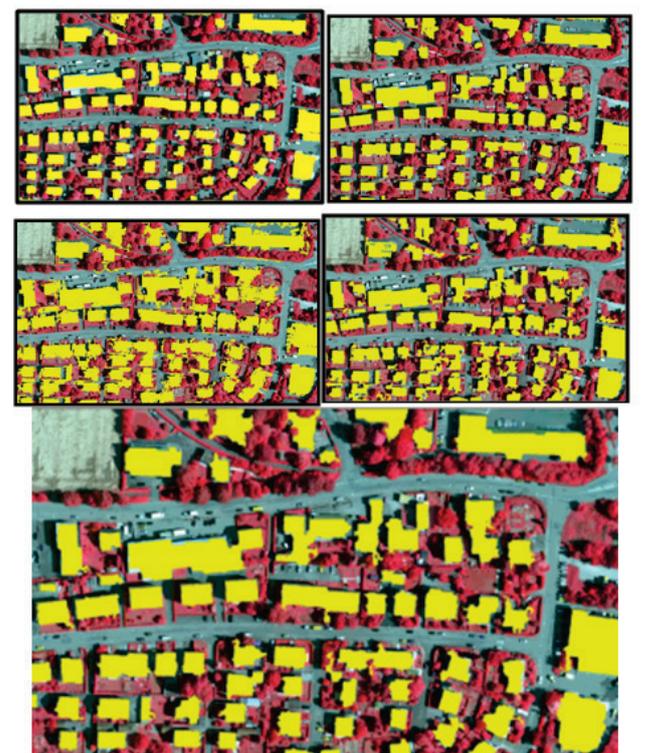
Figure 18 Zoomed view on the residential buildings

Table 5 Quality values of the results from combination of the results

Correctness (%)	94
Completeness (%)	88
Number of objects in the building class	146
Detected objects which intersect with buildings in the reference data	139
Number of buildings in the reference data	218
Number of buildings in the reference data, which don't intersect with any detected object	5

As shown in Tab. 5, the correctness and completeness values are significantly increased by the combination. Remaining problems exist with shadows that are detected as buildings on the vegetated regions. Fig. 18 shows zoomed view on residential buildings. First four images show the results from variants 1, 2, 3 and 4, respectively. The big figure on bottom, corresponds to the final result.

4   Conclusions and discussion

Regarding input data, image and LIDAR data are used in this work. However, the proposed methods, with elimination or slight modifications of some parts, could be used with only one of these input data. Due to complementarities of these data, if possible both should be used. This also depends on financial and other issues but also on the type of the available LIDAR data (e.g. in some countries only post-processed LIDAR DSMs and DTMs are available, in Switzerland, one could use raw DTM data with no or lower point density at buildings and trees respectively, or the whole point cloud). DSMs and DTMs (from which also nDSMs can be produced) play an important role. Especially, their spatial resolution (average point distance), single point accuracy, modelling of surface discontinuities, error characteristics (e.g. amount and size of gross errors), as well as for LIDAR data the ability to use the vertical distribution of semi-transparent surfaces like trees, play a significant role. Of these factors the spatial resolution (GSD of images and average point distance of LIDAR points) is the most important. An improvement of point cloud filtering to derive a DTM, especially for DSM data from image matching, would result in better nDSMs, which also include buildings. Spectral information is important and is currently limited in airborne and most high spatial resolution satellite digital optical sensors to four spectral bands (Red, Green, Blue, Near Infrared). When using classification techniques, an important issue is what other derived channels could be generated with good learning techniques and used for building detection. When using orthoimages, an important issue is their generation with a good DSM to avoid building displacements and roof edge deformation. If possible, even true orthoimages could be used to avoid occlusions. For building detection, buildings are identified without a need of a precise building outline. Four variants of building detection and a combination of them are investigated. In the first variant, a user interaction is required to collect training samples. One of the advantages is that shadows can be detected. The errors in the DSM and DTM cause problems, especially at discontinuities like building outlines. Additionally, green vegetation and green roofs are confused. The second variant is fully automated with the main error sources similar as in the first one. Another drawback of the second variant is that it assumes that all non-vegetated nDSM objects are buildings. Since there are only two classes in the unsupervised NDVI classification, shadows cause errors and may be detected as buildings, even if they are on trees. In the third and fourth variants of building detection, the quality of the LIDAR DTM has a direct influence on the results. In the third variant, the vegetation is again eliminated using NDVI, causing the same problems as the first and second variants due to mixing of green roofs and green vegetation. The main strength of the fourth variant is that it does not require any image data and reduces the problem of separating green vegetation from green roofs and shadow areas on roofs. The quality of the results depends directly on the density of the LIDAR dataset and the amount of echoes per pulse registered. One problem is that walls and other vertical objects close to buildings with a high vertical point distribution are detected as trees, thus some building regions close to the roof outline are not detected. In an attempt to exploit their complementary strengths, an empirical combination of the results of each individual method using intersection and union operations is used. In the evaluation, the combination of the four individual results yields 94 % correct detections and an omission error of 12 %.

The proposed methods detect buildings with no prior information, except the training samples, which are used in the first variant of building detection. Existing datasets, e.g. building maps and cadastral plans that may exist in some countries, are not needed. On the other hand, such information, if of sufficient quality, could improve the results, at least by providing approximations and reducing the search space. The combination of image and LIDAR data provides complementary information, uses mutual advantages and improves the results. The combination of several detection methods provides better and more reliable results. DSM, DTM and nDSM quality have a critical influence on the results. Thus, the goodness of the input data, as well as post-processing methods, mainly DSM filtering for DTM generation, play a critical role and could be improved, starting from both raw LIDAR data or a matching DSM. Major problems in building detection are buildings very close to each other, shadows, occlusions, small and non-planar objects on roofs, vegetation on roofs, trees close to or overlapping roofs, and small roof planes. All these issues need further improvement.

In the detection, the input data of the four variants could be combined in one single method or at least the results of the four variants could have a quality factor for each pixel (e.g. probability of being a building) and thus a weighted combination of the results instead of binary intersections and unions could lead to a better quality.

5　References

[1] Henricsson, O.; Baltsavias, E. 3-D building reconstruction with ARUBA: a qualitative and quantitative evaluation. // A. Gruen, E. P. Baltsavias, O. Henricsson (Eds.), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag, Basel, 1997, pp. 65-76.

[2] Baillard, C.; Zisserman, A. Automatic reconstruction of piecewise planar models from multiple views. // Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition / Colorado, USA, June 23-25, 1999. pp. 559-565.

[3] Koc San, D. Approaches for automatic urban building extraction and updating from high resolution satellite imagery. PhD Thesis, Document No. 238726, The Graduate School of Natural and Applied Sciences, Middle East Technical University, 2009.

[4] Maas, H.-G.; Vosselman, G. Two algorithms for extracting building models from raw laser altimetry data. // ISPRS Journal of Photogrammetry & Remote Sensing. 54, 2-3(1999), pp. 153-163.

[5] Rottensteiner, F. Roof plane segmentation by combining multiple images and point clouds. // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38, Part 3A, 2010, pp. 245-250.

[6] Elberink, S. O.; Vosselman, G. Building reconstruction by target based graph matching on incomplete laser data: analysis and limitations. // Sensors. 9, 8(2009), pp. 6101-6118.

[7] Dorninger, P.; Pfeifer, N. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. // Sensors. 8, 11(2008), pp. 7323-7343.

[8] Baltsavias, E. P. A comparison between photogrammetry and laser scanning. // ISPRS Journal of Photogrammetry and Remote Sensing. 54, 2-3(1999), pp. 83-94.

[9] Rottensteiner, F.; Trinder, J.; Clode, S.; Kubik, K. Automated delineation of roof planes from LIDAR data. // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 36, Part 3-W19, 2005. pp. 221-226.

[10] Sohn, G.; Dowman, I. J. Data fusion of high-resolution satellite imagery and LIDAR data for automatic building extraction. // ISPRS Journal of Photogrammetry and Remote Sensing. 62, 1(2007), pp. 43-63.

[11] Kim, C.; Habib, A. Object-based integration of photogrammetric and LIDAR data for automated generation of complex polyhedral building models. // Sensors. 9, 7(2009), pp. 5679-5701.

[12] Demir, N.; Baltsavias, E. Automated modeling of 3d building roofs using image and LIDAR data. // ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., I-4, 2012, pp 35-40.

[13] Haala, N.; Hastedt, H.; Wolff, K.; Ressl, C.; Baltrusch, S. Digital photogrammetric camera evaluation – Generation of digital elevation models. // Photogrammetrie - Fernerkundung – Geoinformation. 2(2010), pp. 99-115.

[14] Zhang, L. Automatic digital surface model (DSM) generation from linear array images. // PhD Thesis, IGP Mitteilung No. 90, Institute of Geodesy and Photogrammetry, Eidgenössische Technische Hochschule (ETH), Zürich, 2005.

[15] Zhang, K.; Chen, S.; Whitman, D.; Shyu, M.; Yan, J.; Zhang, C. A progressive morphological filter for removing nonground measurements from airborne LIDAR data. // IEEE Transactions on Geoscience and Remote Sensing. 41, 4(2003), pp. 872-882.

[16] Richards, J. A. Remote sensing digital image analysis, Springer-Verlag, New York, 1993.

[17] Swain, P. H.; Davis, S. M. Remote Sensing: The quantitative approach. McGraw-Hill, New York, 1978.

[18] Kraus, K.; Pfeifer, N. Determination of terrain models in wooded areas with airborne laser scanner data. // ISPRS Journal of Photogrammetry and Remote Sensing. 53, 4(1998), pp. 193-203.

[19] Schenk, T.; Csatho, B. Fusion of LIDAR data and aerial imagery for a more complete surface description. // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34, Part 3A, 2002. pp. 310-317.

[20] Lei, C.; Shuhe, Z.; Wenquan, H.; Yun, L. Building detection in an urban area using LIDAR data and QuickBird imagery. // International Journal of Remote Sensing. 33, 16(2012), pp. 5135-5148.

Author's address

Nusret Demir, Dr. Sc. ETH
ETEN Ar-ge Muhendislik Ltd, Antalya, Turkey
Previously at ETH Zurich Institute of Geodesy and Photogrammetry
nusret.demir@eten.com.tr