



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Automated mapping of buildings through classification of DSM-based ortho-images and cartographic enhancement

Höhle, Joachim

Published in:

International Journal of Applied Earth Observation and Geoinformation

DOI (link to publication from Publisher):

[10.1016/j.jag.2020.102237](https://doi.org/10.1016/j.jag.2020.102237)

Creative Commons License

CC BY-NC-ND 4.0

Publication date:

2021

Document Version

Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Höhle, J. (2021). Automated mapping of buildings through classification of DSM-based ortho-images and cartographic enhancement. *International Journal of Applied Earth Observation and Geoinformation*, 95(3), Article 102237. <https://doi.org/10.1016/j.jag.2020.102237>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

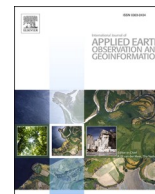
Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Automated mapping of buildings through classification of DSM-based ortho-images and cartographic enhancement

Joachim Höhle

Aalborg University, Department of Planning, Rendsburggade 14, 9000 Aalborg, Denmark

ARTICLE INFO

Keywords:

Mapping
Map updating
Automation
Classification
Regularization
GIS

ABSTRACT

Urban areas are changing rapidly. In order to document the urban realities in topographic databases and geographic information systems efficient methods are required. Vector data of buildings are of special importance. A methodology for an automated generation of cartographically enhanced data is presented and applied to two test sites at Vaihingen, Germany. The steps of the workflow are described in detail. The examples use imagery of a large-format aerial camera to map different types of buildings. First, land cover maps are generated by means of supervised classification using two sets of attributes (basic attributes and attribute profile, basic attributes and dispersion). After the enhancement of the extracted buildings their outlines have straight, orthogonal, and parallel line segments created by least squares adjustment. The assessment of the geometric accuracy used 264 well-defined building corners and two types of references (land cover map, ortho-image). The obtained average standard deviation of the coordinates was $\sigma_{x,y} = 1.0$ m. The additional use of an attribute profile did not improve upon the geometric accuracy that was obtained by means of five attributes (height above ground, normalized difference vegetation index, standard deviation of the elevations in the 5×5 pixels window, intensity value of the near-infrared band, and standard deviation of intensities in the 5×5 surrounding at a pixel of the near-infrared band). The experiences with the developed software reveal that a graphical output of intermediate results is helpful to obtain complete and reliable results at complex building structures.

1. Introduction

Mapping and map updating of built-up areas require accurate geodata in vector format. At present, such data are mainly produced by highly skilled operators using expensive stereo-workstations (Spreckels et al., 2010). There have been many studies to make this task faster and cheaper (Mayer, 1999; Oude Elberink, 2008). There is a not insignificant need to increase the map production in the world. Only 30 % of the land area is mapped in scales of 1:25000 or greater. The time interval for updating these maps is between 10 and 30 years in some countries (Konecny et al., 2016). New attempts to supply topographic data are supported by Google and Microsoft (Google maps, Bing maps). The OpenStreetMap is a voluntary crowd sourcing attempt to update national topographic maps by the public (OSM Wiki, 2020). Private firms use imagery of small-format cameras installed in drones for mapping of small areas (Mayr, 2011; He et al., 2019). All these tasks require efficient methods and they should preferably be automatic. This article seeks to contribute to these attempts. The development of new cameras, image processing and machine learning methods gives hope to achieve this

goal. The automated generation of land cover maps (LCM) and land use maps has benefitted from these new tools (Maxwell et al., 2018). This study uses large-format aerial camera imagery as source data for a new approach in the automated 2D mapping of buildings. Despite increasing use of both laser scanning and very high spatial resolution satellite data, large-format aerial camera imagery is still widely collected on a routine bases and applied for map updating (Heipke et al., 2008).

The requirements in mapping of buildings differ between contexts. The objects to be discussed in this contribution are the outlines of the building roofs. They differ in position from the building walls. For the latter, ground surveying must be applied to determine the roof overhangs. This is usually not done for the core data of topographic data bases at scales between 1:5000 to 1:10000. The requirements for the planimetric accuracy of building roofs are a few decimetres only. Other requirements regard the completeness of the data. The minimum size of the building area and of its sides to be mapped may also be part of the specification. If a sufficiently small ground sampling distance (GSD) is selected for the image acquisition, all these cartographic requirements should be met by the automated process. For map updating, the use of

E-mail address: jh@plan.aau.dk.

<https://doi.org/10.1016/j.jag.2020.102237>

Received 27 October 2019; Received in revised form 29 July 2020; Accepted 2 September 2020

Available online 6 October 2020

0303-2434/© 2020 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

planning data is practised in some countries. This may be more economic, but the homogeneity of the data will suffer since planning data update is often piecemeal. Thus, airborne survey methods are widely used for new mapping of large areas.

Relevant automated method development has been carried out for several years including for the two major steps of the proposed solution, classification for building extraction and enhancement (regularization) of the roof outlines. In recent years, classification of images has received many innovations from the field of image analysis and machine learning. The ISPRS III/4 working group has provided ortho-images and digital elevation models for validation of results of several research studies using these data, i.e. the “2D semantic labelling contest” (ISPRS WG III/4, 2014). That initiative has seen some notable successes regarding the thematic accuracy of topographic and non-topographic objects, for example, thematic accuracy of buildings of above 95 %, was achieved by Marmanis et al. (2018). The results in Damadoran et al. (2017) revealed also remarkably high thematic accuracies for buildings using a variety of attributes and machine learning techniques at another test site. The applied attributes were a combination of basic attributes, i.e., intensities (I) of the ortho-image, digital surface model (DSM), and normalized difference vegetation index (NDVI), with attribute profiles. The latter were based on DSM- and NDVI-values.

Regarding the cartographic enhancement of buildings, recent research has been carried out by Wang (2016); Tasar et al. (2018), and Mousa et al. (2019). The approach of (Wang, 2016) generates the outlines of buildings from a dense point cloud which is derived by image matching. Buildings are extracted by thresholding the gradients of elevations. Building roofs are then classified using radiometric features. The edge pixels of the roof are traced, and closed polygons are generated. By a split-and-merge process segments are obtained, and straight lines are fitted to them. Successive lines are intersected, and roof corners are obtained. 99 % of the detached buildings were detected. Geometric accuracies, i.e. positional errors of building outlines, were not reported.

In Tasar et al. (2018) a fine mesh is placed on top of the classified map. Labelled triangle meshes are then optimized by means of an objective function, which balances the closeness to the classification map, the rectangularity of the building edges, and the mesh complexity. The approach yields an 80 % overlap of the processed buildings to their reference. Small building edges can be detected and mapped. The generated building edges are not perfectly straight, orthogonal, and parallel. Again, measures of the geometric accuracy are not reported. The approach of Mousa et al. (2019) starts the enhancement of building outlines from a gridded DSM. An algorithm is used which finds corner points by means of a likelihood function. A simplification procedure is applied which uses different building models (rectangle and rectilinear polygons). The parameters of the rectilinear polygon (orientation angle of the lines and distances of the line segments from origin) are found by least squares adjustment. The geometric accuracy of the derived vertices is quoted with $RMSE_{dp} = 0.9$ m for the ISPRS test data “Vaihingen”. The applied accuracy measure (dp) uses the perpendicular distance between each corner (vertex) of an extracted building polygon and the nearest boundary point of the reference polygon. Errors above 3 m were excluded from the calculation of the accuracy measure.

Other investigations have applied data from Airborne Laser Scanning (ALS) and identify, trace, and regularize the outlines of buildings. In Awrangjeb (2016), e.g., a Delaunay triangulation is applied to an ALS point cloud. A boundary line consisting of small segments can then be identified. The line segments are smoothed, and corners are derived. The achieved average accuracy of the distance between corners (d) for three areas of the ISPRS data set “Vaihingen” is quoted with $RMSE_d = 0.7$ m.

However, it seems that a practical method for the automated mapping of buildings and other topographic objects is still missing. It is the first goal of this study to outline a method for 2D mapping and map updating of buildings. Such a method should be able to produce vector data of high cartographic quality and geometric accuracy. Its practicality should be evaluated for a diverse set of buildings. The used

methodology must be applicable for simple as well as for complex building structures. Furthermore, the detection and extraction of the outline segments must be independent of their direction and length. The automated processing of various building types may produce some errors. An operator should, therefore, have the possibility to detect the reason of the errors and remove them quickly. It is proposed that a graphic output of intermediate results will assist the operator in such editing work. As a benchmark, the achievable planimetric accuracy (σ_x, σ_y) for well-defined points using manual stereophotogrammetry is about 1/3 of the ground sampling distance (Spreckels et al., 2010).

This contribution will focus on the use of large-format aerial imagery. Such images have advantages by their high spatial resolution and availability of four spectral channels of large spectral bandwidth. These characteristics enable the automatic connection of images by conjugate points which results in high planimetric accuracy for the determined objects as well as in the automatic recognition of all types of topographic objects. Elevation data can be generated from overlapping images with extremely high density and accuracy (Haala and Rothermel, 2012). Ortho-images can easily be derived using digital surface models (DSM) in the rectification. Buildings and other objects above terrain can be mapped without displacements when using DSM-based ortho-images. The use of DSM-based ortho-images is therefore a prerequisite for the suggested methodology. Another prerequisite is a high-quality LCM from which all buildings in the ortho-image can be extracted and enhanced. The land cover map may be derived by a classifier using various attributes. Usage of attribute profiles together with some basic attributes is a new approach in the generation of LCM and it is a second goal of this study to obtain experience with that. The LCM generated by a simple attribute combination described in Höhle (2017) will be processed for comparison. The enhancement to straight, parallel, and orthogonal vectors takes place in both examples using the same methodology. To gain experience with the diversity of buildings, two test sites will be processed. A third goal of this study is to improve the software package for the automated generation of building vectors from DSM-based ortho-images with high cartographic quality and geometric accuracy.

2. Materials and methods

2.1. Materials

A data set of ISPRS (ISPRS WG III/4, 2014) was used, which includes several DSM-based ortho-images and land cover maps. In addition, a data set of normalized digital surface models (nDSM) of the same area was available (Gerke, 2014).

The provided digital surface model has been generated from imagery using dense image matching (Haala and Rothermel, 2012). Its grid size is 0.09 m. The ortho-images were produced by using the same DSMs for the differential rectification of the aerial images. The false colour ortho-images have a pixel size of 0.09 m and three spectral channels (Near Infra-Red, Red, and Green). The ortho-images are produced from the aerial images of high overlaps (65 % in flight direction, 60 % across flight direction). In the provided data, these are referred to as true ortho-images. According to current Danish specification, true ortho-images are produced by using a detailed 3D-building model (GeoForum, 2011). In Germany, the production of the so-called “True-DOP” uses a digital surface model which is derived by dense image matching (Baltrusch, 2016). In order to avoid confusion, the provided ortho-images are here named ‘DSM-based ortho-images’ in contrast to ortho-images which use 3D-building models or digital terrain models (DTM).

As the DSM-based ortho-image has no displacements at objects above ground, the use of DSM-based ortho-images is a key feature of the proposed method for the automated generation of vector data for buildings. The provided nDSM has been derived by classification of the DSM into two classes (“ground”, “above ground”) and subsequent calculation of

the difference in elevation (Axelsson, 2000). The ISPRS provided land cover maps are produced by manual digitization and are converted into raster format.

Two test sites were selected. They differ in the building characteristics (density, number of corners, overlapping trees). At Example 1 (Vaihingen, area 7), the buildings in the 2.8 ha large area are quite varied. In the upper part of the area, the buildings are large and detached. Tall trees and bushes are close to the buildings. In the lower part, the buildings are of complex shape and close to each other. There are only a few trees. Buildings and trees had long shadows at the time of imaging. The landscape of Example 2 (Vaihingen, area 1) contains 28 buildings of complex shape and high density. Tall trees partially overlap the buildings. This test site has an area of 3.9 ha.

For the assessment of the thematic and geometric accuracy, the

provided land cover maps and the ortho-images are used. The assessment of the thematic accuracy of a derived land cover map may be done by means of a subset of pixels or by all pixels of the reference data. Well-defined corners of buildings were used to assess the geometric accuracy.

2.2. Methods

2.2.1. Overview

The proposed methodology uses overlapping aerial images from which a digital surface model (DSM), a DSM-based ortho-image, and various attributes will be derived. In this study, two different methods are applied for the generation of LCM: for Example 1, supervised classification using 17 attributes, and for Example 2, supervised classification using just five attributes. Both land cover maps are generated using

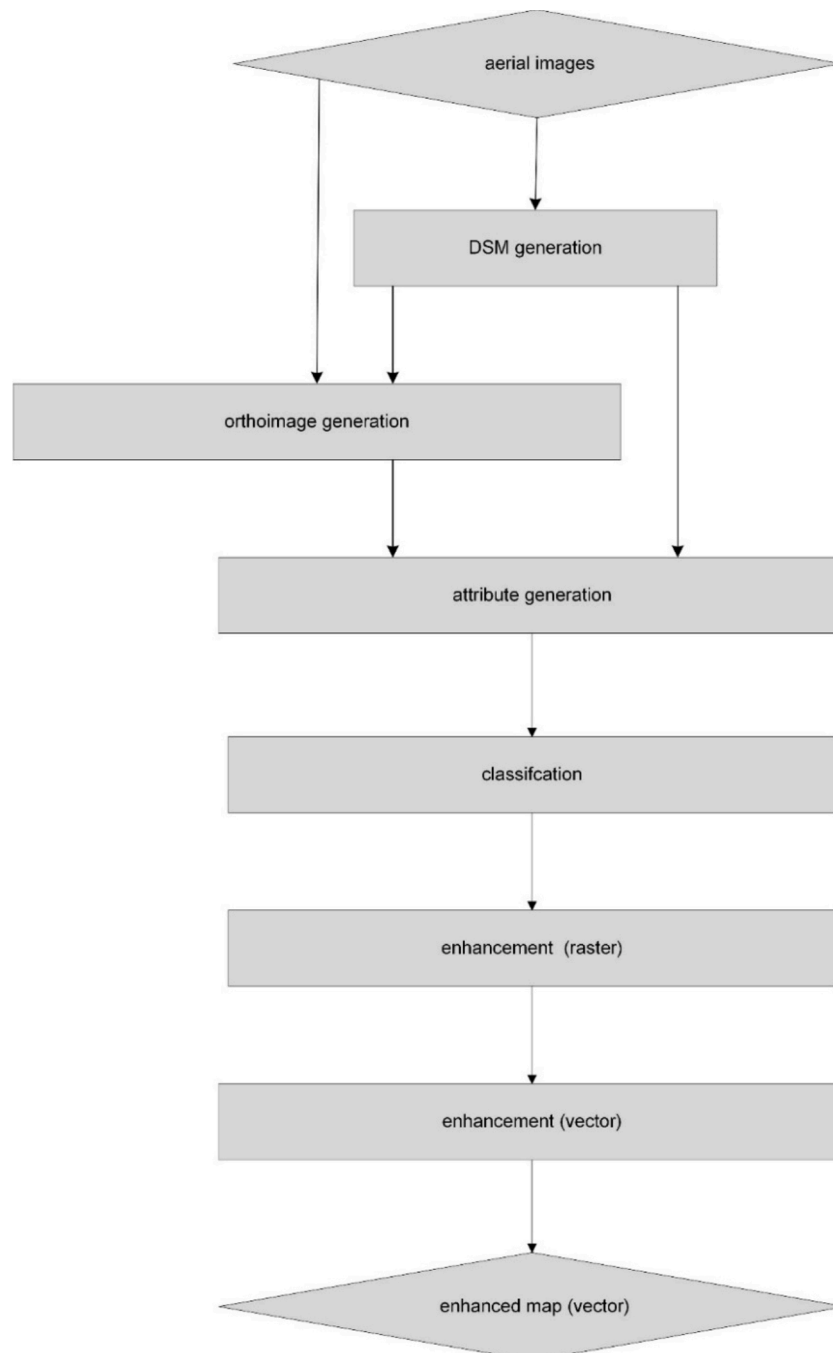


Fig. 1. Flow chart of the steps in the applied methodology.

the Decision Tree (DT) classifier. They are named LCM_1 (Example 1) and LCM_2 (Example 2). It is an advantage of the DT classifier, that categorical and numerical data can be used and that a normal distribution of the variables is not required (Kamusoko, 2019). The theoretical background of the DT method is given in Breiman et al. (1984). Experiences with DT classification for the generation of land cover maps are reported in Friedl and Brodley (1997) and Höhle (2014).

The class “building” is extracted and then cartographically enhanced. The final building outlines are generated by means of line detection, line sequence determination, and least squares adjustment. The flow chart in Fig. 1 depicts the various steps of the applied methodology.

2.2.2. Attribute generation

Attributes of the pixels are used to generate the land cover maps. The applied attributes must characterize the selected classes and make it possible to distinguish the classes of the land cover map from one another. Basic attributes are the intensities of the three spectral bands, the normalized difference vegetation index, and the height above ground. A new approach used by this study, for Example 1, is an attribute profile (AP). The AP consists of several filtered images which are generated from the bands of the ortho-image. For each of the bands a tree of connected components (CC_AP) is created and then pruned using thresholds for an attribute of the CC_AP at different intensity levels. For example, a ‘maxTree’ - algorithm can be applied where the filtering of the created CC_AP is carried out using the attribute “area” for thresholding. The pruned pixels are re-allocated to the CC_AP of the node with a lower level of intensity. When three bands and three thresholds are selected, nine additional intensity values in the vector of each pixel are created. Together with the intensities of the original three bands, this multi-scale Area_AP comprises $4 \times 3 = 12$ intensity values.

Such filtering of the image enables that small bright regions are merged into larger darker regions. This procedure is named thinning attribute filtering (Dalla Mura et al., 2010; Damadoran et al., 2017). Fig. 2 illustrates the principle of this filtering.

Also, a thickening attribute filtering could be created when using the ‘minTree’ - algorithm, which increases the number of features by 12. In this study, only the thinning attribute filtering is used and combined with basic attributes, i.e. I_{NIR} , I_R , I_G , height above ground, normalized difference vegetation index at Example 1. It is anticipated that the selected filtering removes the small, light areas within the mostly dark roofs and thus enables a better detection of buildings. In addition, the feature vector should be manageable by small computers.

Other special attributes used for Example 2 are the standard deviations of elevations (Z) and of intensities in the 5×5 pixel kernel surrounding of a pixel (dispersion). They are combined with three basic attributes (I_{NIR} , nDSM, NDVI). The use of the two attributes (I_{NIR} , sigma

$I_{NIR_{5 \times 5}}$) considers that the NIR-band has large changes in intensity for vegetation. Trees may then be distinguished from buildings. The other available bands (Red, Green) are omitted in the model to keep the number of features at a minimum. These attributes were applied in Höhle (2017) to create the land cover map (LCM_2), which is used in this study to extract and enhance many buildings.

All selected attributes are collected in a vector, which is named a “feature vector”. Each pixel of the new land cover map is represented by a feature vector. The attributes used in the two examples are presented in Table 1.

For Example 1, three thresholds ($T_{area} = 100, 1000, 5000$ pixels) are applied. The special attributes (Area_AP) comprise 12 attributes and the basic attributes comprise five attributes. Since the buildings in the used ortho-image all have dark roofs, the roof areas become more homogeneous using small area thresholds (100 and 1000 pixels). The filtering with a 5000 pixels large area threshold will cause that buildings smaller than $6.4m \times 6.4m$ are not recognizable in the filtered image.

2.2.3. Classification

There are two steps involved in classification: the generation of the classifier and the assignment of a class to each map unit. As a classifier, the “Decision Tree” (DT) is selected for both data sets. The derivation of the DT requires training data, which were available in this investigation through a raster land cover map, in which the pixel value represents a single class. Pixels, either for a sub-map or the whole map, are extracted

Table 1
Used attributes in the two examples/methods to generate LCMs.

Attribute type	Attributes	
	Example/Method	
	1	2
Basic	I_{NIR}	I_{NIR}
	I_R	–
	I_G	–
	NDVI	NDVI
	nDSM	nDSM
Special	I_{NIR}	sigma $Z_{5 \times 5}$
	$I_{NIR_{100}}$	sigma $I_{NIR_{5 \times 5}}$
	$I_{NIR_{1000}}$	–
	$I_{NIR_{5000}}$	–
	I_R	–
	$I_{R_{100}}$	–
	$I_{R_{1000}}$	–
	$I_{R_{5000}}$	–
	I_G	–
	$I_{G_{100}}$	–
	$I_{G_{1000}}$	–
	$I_{G_{5000}}$	–

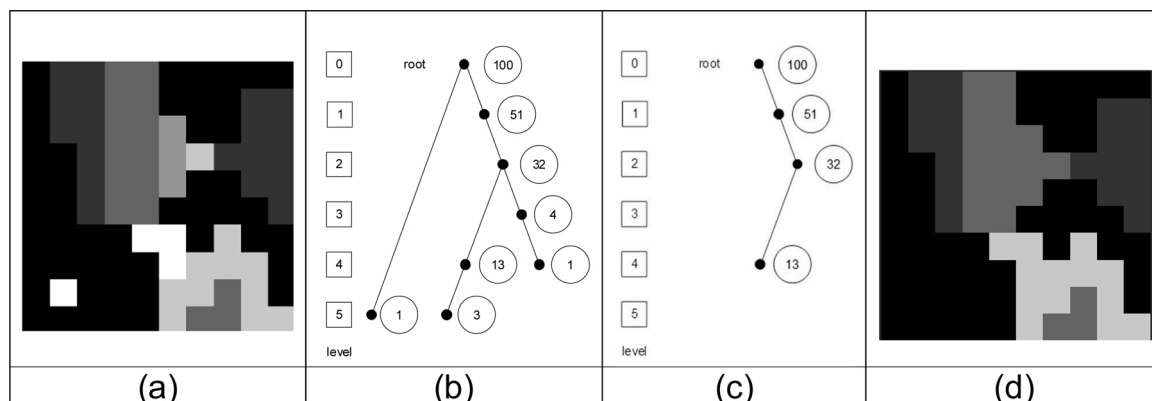


Fig. 2. Principle of thinning attribute filtering. Original image (a), maxTree with area size and levels (b), pruned tree after applying the threshold $T_{area} > 5$ pixels (c), and attribute (“area”) filtered image (d).

and supplemented with attributes. The DT classifier splits the training data so that an optimal homogeneity of a class within a subset is achieved. This splitting of the data occurs several times until all training units (pixels) are separated into the selected classes. The tree is then pruned. By means of the derived DT, the class of each map pixel can be predicted. In the examples, five and six classes are selected. The training of the classifier (DT) was carried out with 500 samples per class at Example 1. At Example 2, all pixels of an adjacent land cover map (Vaihingen, area 26) consisting of 5.3 million pixels were used. In this way, the training data are independent of the test site data and all classes of the LCM can be trained.

Since the quality of the derived land cover map influences the final vector map, an assessment of the thematic accuracy must be carried out. The derived class (category) for each pixel is compared with the “truth”, i.e., the class value of the corresponding pixel of the provided land cover map (reference). For this purpose, an error matrix is established, and accuracy measures are derived as suggested by Congalton and Green (2008).

2.2.4. Cartographic enhancement

The result of the classification is a land cover map of several classes, in raster format. The class “building” is extracted, enhanced, and the connected components (CC_Enh) are labelled using standard image processing methods (dilation, erosion, edge detection, segmentation, hull filling, feature extraction). Corresponding functions are contained in the open-source software package “EBImage” (cf. Section 4.2).

The detection of line segments in the CC_Enh representing one building uses the Hough transformation where a line is expressed by Eq. (1).

$$\rho = x \cos \theta + y \sin \theta \tag{1}$$

where

- ρ =orthogonal distance of the line from the origin of the xy -system
- θ =azimuth of the normal vector to the line
- x, y =pixel coordinates.

The coordinates (x, y) are constants in the parameter space $H(\theta, \rho)$.

The pixels of the CC_Enh are transformed into the parameter space using several θ -values in the range between 0° and 175° . All pixels belonging to a line segment are accumulated in cells of the parameter space. Its resolution is 5° and 5 pixels, respectively. The cell with the maximum number of counts is considered as the reference line $(\theta_{ref}, \rho_{ref})$. The other lines are found by analysing the result of the Hough transform. The applied detection criteria are ρ_{min} , ρ_{max} , and n (length of line) for θ_{ref} and $\theta_{ref} + 90^\circ$.

A graphical output may support the detection of all segments of the building outline. When all segments of the outline are identified, all pixels of the cluster representing a line are extracted. Accurate line parameters are determined by least squares adjustment. The applied method is based on Eq. (1). All residuals are then orthogonal to the line. The sequence of lines is determined by means of the angle between the centre of the CC_Enh and the centre of the cluster which represents a line segment. Successive lines are then intersected, and approximate corner positions are obtained by Eq. (2)

$$\begin{pmatrix} x_{p_k} \\ y_{p_k} \end{pmatrix} = \begin{pmatrix} q_1 & q_2 \\ q_3 & q_4 \end{pmatrix} \begin{pmatrix} \rho_i \\ \rho_{i+1} \end{pmatrix} \tag{2}$$

where

- x_p, y_p = coordinates of an intersected corner point,
- q_i = coefficients containing trigonometric functions of successive angles θ ,
- ρ_i = orthogonal distance of the line from the origin.

Accurate coordinates of the corner points are calculated in two steps. First, the weighted average of the angle (θ_{av}) is determined whereby the weights are derived from the lengths of the lines. A least squares adjustment derives new values for the orthogonal distances of the lines

from the origin (ρ_i). The new values for the corner coordinates (p_i) can then be derived by multiplying the design matrix (A) with the calculated unknowns ($x_{i=\rho_i}$).

$$p = A \hat{x} \tag{3}$$

The accuracy for a building with n corners can be estimated by the standard deviation of the residuals (r).

$$\hat{\sigma}_r = \sqrt{\frac{\hat{r}^T \hat{r}}{n}} \tag{4}$$

All line segments of the building are then either parallel or orthogonal. The standard deviation of the residuals is a measure of the interior accuracy. When a residual exceeds $T_r = 3\sigma_r$, a warning is displayed. The polygon of each building is closed. The processing of the raw outlines into vectors is depicted in Fig. 3.

More details on the mathematics of the used approach for deriving orthogonal and parallel line segments are given in Höhle (2017). Some improvements regarding the universal use are introduced in the new version of the developed software (cf. Section 4.2). The adjustment of the lines is now based on Eq. (1). The residuals are then orthogonal to the line. Lines that are almost parallel to the y -axis can now be adjusted. In addition, the pixel cluster representing the line is cleaned of pixels belonging to another line. This is achieved by analysing histograms. The calculated line parameters are then more precise.

2.3. Assessment

The automated mapping of building outlines by means of vectors is evaluated in two ways. A visual inspection of the generated building outlines together with the references (land cover map, the DSM-based ortho-image) will give a first impression whether the extraction of buildings and their enhancement have been successful. To access the geometric accuracy of the derived buildings, check points must be selected. They should be well-defined in both references and in the generated vector data. Suitable check points are the corners of buildings. Their intersecting lines should form an angle of about 90° . Reference coordinates were determined by manual digitizing of the selected corners in the two references. The applied accuracy measures are the standard deviation (σ) and the root mean square error (RMSE). They are determined for both coordinates (x, y) by

$$\sigma = \sqrt{\frac{\sum(\Delta - \mu)^2}{n - 1}} \tag{6}$$

and

$$RMSE = \sqrt{\frac{\sum \Delta^2}{n}} \tag{7}$$

where

- Δ = coordinate difference between reference and enhanced building corner,
- μ = average coordinate difference
- n = number of check points.

The accuracy measures are calculated for each building and the average of the accuracy measure for all processed buildings is cited. Gross errors must be eliminated. They are defined as gross error $> 3\sigma_{av}$ and their number ($n_{gross\ error}$) is quoted in the results. To find the reason for gross errors and inaccuracies, the intermediate results are also of interest, which concerns the thematic accuracy of the class “building” in the classification.

The thematic accuracy is determined pixel-wise. Reference values are extracted from the provided land cover map. For comparison of the classification result with the reference, an error matrix is established (Congalton and Green, 2008). The accuracy measures user accuracy

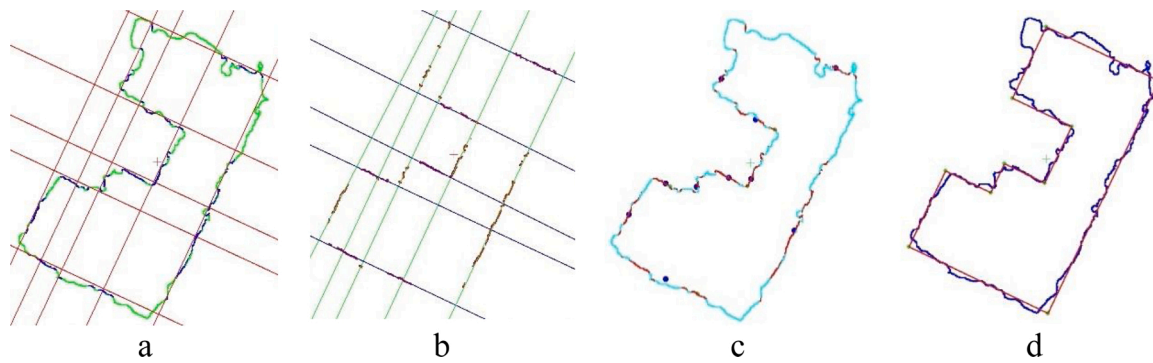


Fig. 3. Processing of an enhanced raster map into vectors (a...detected lines, b...adjusted lines, c...centre of the clusters representing lines, d...adjusted corner points of the building).

(UA), producer accuracy (PA), and the harmonic mean of UA and PA (H) are calculated for the class “building”. Regarding the geometric accuracy of the vector data, the residuals after the least squares adjustment of the building polygon may disclose gross errors in the corner points. The derived average standard deviation ($\sigma_{r,av}$) is a measure for the interior accuracy of the corner points. When the two references are compared with each other, the average standard deviation of the coordinate differences ($\sigma_{\Delta x, \Delta y}$) is a measure for the quality of the references.

3. Results

The derived land cover maps (LCM_1, LCM_2) comprise the classes “building”, “impervious surfaces”, “low vegetation”, “tree”, and “car”. For LCM_2, an additional class (“clutter/background”) was selected to detect water bodies and a few other small objects (tennis courts, containers, swimming pools). For the enhancement, thresholds for the minimum area of a building (25 m²) and minimum length of a line segment (1.4 m) were used. Only buildings which are completely contained in the ortho-image have been processed. Manual editing was not carried out. The results are presented by overlaying the vector data onto the references (land cover map, ortho-image) and by numeric values of an accuracy measure (standard deviation, RMSE).

3.1. Example 1

The thematic accuracy of the derived land cover map (LCM_1) is assessed by means of the provided land cover map consisting of 4.8 million pixels. For the class “building”, it resulted in PA = 92 %, UA = 82 %, and H = 87 %.

A visual inspection of the generated vector map in Fig. 4 reveals a high cartographic quality for all vectors. The straight, parallel, and orthogonal lines are perfectly established. Non-orthogonal lines are also present in the enhanced vector map. Some of the lines are short. The buildings have up to 12 corners. All outlines of the 28 automatically mapped buildings are closed. A few erroneous overlaps between buildings occurred.

The vector map of the buildings is superimposed onto the two references, i.e. the ortho-image and the land cover map, which allows a visual evaluation of the quality of the automated processing. The outlines of the buildings fit well at most of the buildings in the ortho-image (Fig. 5). Single buildings have been combined to form larger units. Two buildings were not detected by the automated processing. Errors occurred when trees overlap the buildings. Shadows of tall buildings and trees are sources of errors.

It is apparent that the enhanced vector data are more generalized than the buildings of the provided land cover map (Fig. 6). A few errors of the enhanced buildings are visible. However, errors exist also in the land cover map. For example, when trees overlap buildings, then the outline of the overlapping tree has been compiled.



Fig. 4. Vector map derived from the DSM-based ortho-image by the developed method (Vaihingen, Germany, area 7).

The assessment of the geometric accuracy used 130 well-defined building corners as check points. Fig. 7 depicts the distribution of the check points along with the buildings which are generated from the provided land cover map (reference).

Geometric accuracy metrics are shown in Table 2. Two references have been used, the land cover map and the DSM-based ortho-image. The results are nearly the same for both references. The average of the standard deviations of the corner coordinates is $\sigma_{x,y,av} = 1.05$ m. 130 or 142 well-defined corners have been used to assess the geometric accuracy. Two gross errors have been removed before the final calculation of the accuracy measure.

The average residual after the adjustment of the corner coordinates were $\sigma_{r,av} = 0.2$ m only. Comparing the two references with each other, the calculated standard deviation of the coordinate differences is $\sigma_{\Delta x} = 0.6$ m and $\sigma_{\Delta y} = 0.7$ m ($n_{corner} = 132$, $n_{gross\ error} = 0$).

3.2. Example 2

The generated land cover map (LCM_2) has six classes. Its thematic accuracy was assessed by reference values for each of the 4.9 million pixels. For the class “building” the selected accuracy measures were PA



Fig. 5. Enhanced vector data of buildings superimposed onto the DSM-based ortho-image. The green dots mark undetected buildings. Source of ortho-image: ISPRS WG III/4 (2014).

= 78 %, UA = 88 %, and H = 83 % (Höhle, 2017).

The visual inspection of the generated vector data (Fig. 8) reveals a high cartographic quality for all buildings. The outlines of their roofs include orthogonal and non-orthogonal lines. Some of the lines are short. The buildings have up to 14 corners. All polygons of the 33 buildings are closed.

Fig. 9 depicts the derived buildings together with the ortho-image. The outlines of the buildings fit well at most of the buildings in the ortho-image. One building was not detected. Some single buildings are combined to larger units. Shadows of tall buildings have been a source of errors.

Fig. 10 depicts the enhanced vector data together with the provided land cover map (reference). As for Example 1, it is apparent that the enhanced vector data are more generalized than the buildings of the land cover map. A few errors in the enhanced vector map are visible. For example, a flat part of the large unit on the lower left side of the map is missing. The human operator responsible for production of the reference map may have used the shadows to recognize and map the building.

The assessment of the geometric accuracy used a high number of well-defined building corners as check points. Fig. 11 depicts the distribution of the check points along with the buildings which are generated from the provided land cover map (reference).

Numerical values for the geometric accuracy of the processed buildings are listed in Table 3. The results are nearly the same for both references. The average of the standard deviation of the corner coordinates is $\sigma_{x,y,av} = 0.9$ m. The results are based on 134 (154) well-defined corners. One gross error has been removed before the calculation of the accuracy measure.

The average standard deviation of the residuals after the least squares adjustment of the corner

coordinates were $\sigma_{r,av} = 0.3$ m. When the two references are compared, the calculated standard deviation of the coordinate differences is $\sigma_{\Delta x} = 0.3$ m and $\sigma_{\Delta y} = 0.3$ m ($n_{corner} = 130$, $n_{gross\ error} = 2$).

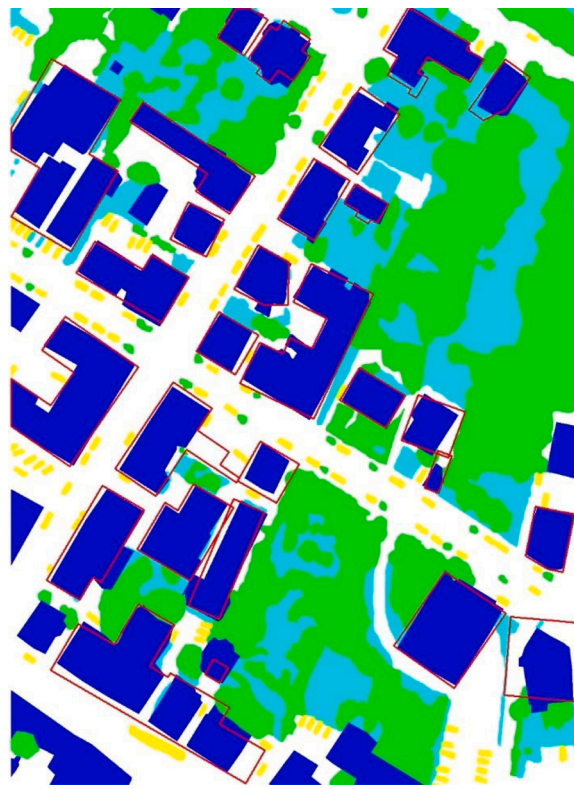


Fig. 6. Enhanced buildings (red lines) superimposed onto the manually compiled land cover map (reference). The categories of the land cover map are coded by colours: “building” (blue), “impervious surfaces” (white), “low vegetation” (cyan), “tree” (green), “car” (yellow). Source of land cover map: ISPRS WG III/4 (2014).

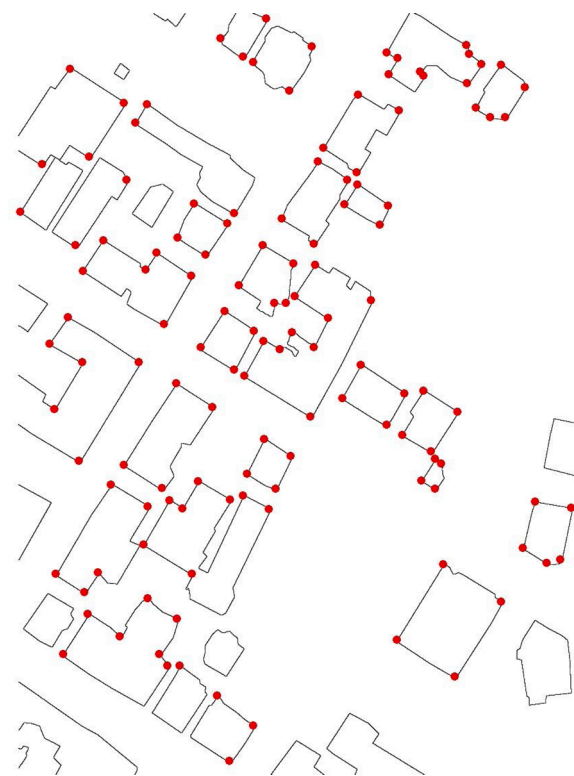


Fig. 7. Distribution of the check points along with the buildings generated from the provided land cover map (Vaihingen, Germany, area 7).

Table 2
Geometric accuracy of derived corner coordinates at Example 1.

reference	land cover map		ortho-image	
	x	y	x	y
coordinate				
σ [m]	1.1	0.9	1.2	1.0
n_{corner}	130		142	
$n_{\text{gross error}}$	2		2	

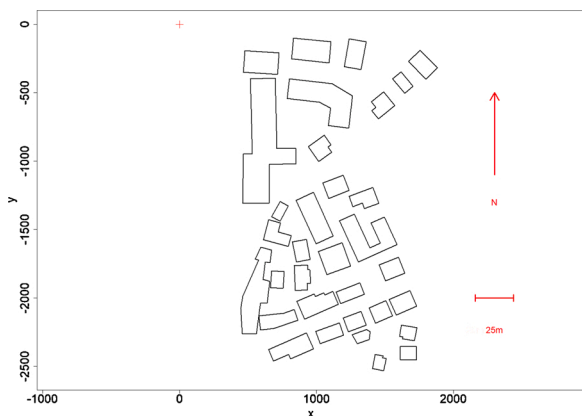


Fig. 8. Vector map derived from a DSM-based ortho-image by the developed method (Vaihingen, Germany, area 1).

4. Discussion

4.1. Vector maps

Visual inspection of the automatically generated vector maps (cf. Figs. 4 and 8) reveals a high cartographic quality. The visual comparison with the two references, however, displays some inaccuracies. Some buildings are not mapped or do not overlap completely. Some deficiencies in the main orientation angle of the buildings are noticeable. The human eye is extremely sensitive to such deviations. The geometric accuracy of the two examples is about the same. The average of the standard deviation is $\sigma_{x,y,av} = 1.0$ m (reference = LCM) and $\sigma_{x,y,av} = 1.0$ m (reference = ortho-image). These accuracies are determined by means of 264 and 296 well-defined corners, respectively.

Of interest is the influence of the applied methodology. Example 1 used for the generation of the land cover map (LCM_1) 17 attributes including an attribute profile, Example 2 used five attributes only to produce LCM_2. The differences in the geometric accuracy of the building vectors are $\Delta\sigma_{x,y} = 0.05$ m (reference = LCM) and $\Delta\sigma_{x,y} = 0.25$ m (reference = ortho-image), both in favour of Example 2. That indicates that the use of an area attribute profile could not improve the geometric accuracy obtained with the five attributes. This means that computers with less main memory (RAM), e.g. just eight GB, can be used for the computations of the LCM and of the building vectors. The comparison of both references reveals that the average standard deviations of the differences are relatively high ($\sigma_{\Delta x, \Delta y, av} = 0.5$ m, $n_{\text{corner}} = 262$, $n_{\text{gross error}} = 2$). This means that the limited accuracy of the references has an influence on the geometric accuracy. Taking this fact into account, the final accuracy of the generated building corners must be better than the calculated values. The quoted accuracy is an absolute accuracy because the used ortho-images are based on a DSM. The reduced image quality of the DSM-based ortho-images creates problems at the detection of lines and their enhancement. The imaged outlines of buildings are not straight and uninterrupted lines in the DSM-based ortho-image.

Both test sites are difficult for an automated system. Some manual interaction was necessary. Better conditions exist for detached buildings without overlapping trees. The harmful influence of vegetation and



Fig. 9. Generated vector data of buildings superimposed onto the DSM-based ortho-image. The green dot marks an undetected building. Source of ortho-image: ISPRS WG III/4 (2014).

shadows can be reduced by proper flight planning. For mapping purposes, the images are usually taken before foliation and at a high position of the sun. For example, according to specifications currently applied in Denmark, the imagery for mapping and ortho-imaging must be taken when the sun is more than 25° above the horizon (Geoforum, 2011).

A comparison with related recent work may give an idea about the performance of this work. Table 4 displays the geometric accuracy of building outlines achieved by different authors. The applied accuracy measure is the root mean square error (RMSE). The definition of the errors is not the same in the different works. Mousa et al. (2019) used the perpendicular distance (dp) to a line segment, Awrangjeb (2016) applied the distance between corresponding building corners and this study used the average of both coordinates (x,y) of the reference system. Other differences concern the way gross errors are dealt with. A fixed value can be selected as a threshold value to eliminate gross errors. A threshold value can also be calculated using a formula, and the threshold value is then a variable.

The source data used by the authors listed in Table 4 are either only airborne laser scanning (ALS), only aerial images, or a combination of ALS and aerial images. The results of Awrangjeb (2016) are the best of the three tests because the distance error (RMSE_d) is always larger than the coordinate error (RMSE_{x,y}).

The results of this study (RMSE_{x,y,av} = 1.0 m, $n_{\text{corner}} = 264$, $n_{\text{gross error}} = 2$, reference = LCM) are nearly the same as the results reported in}

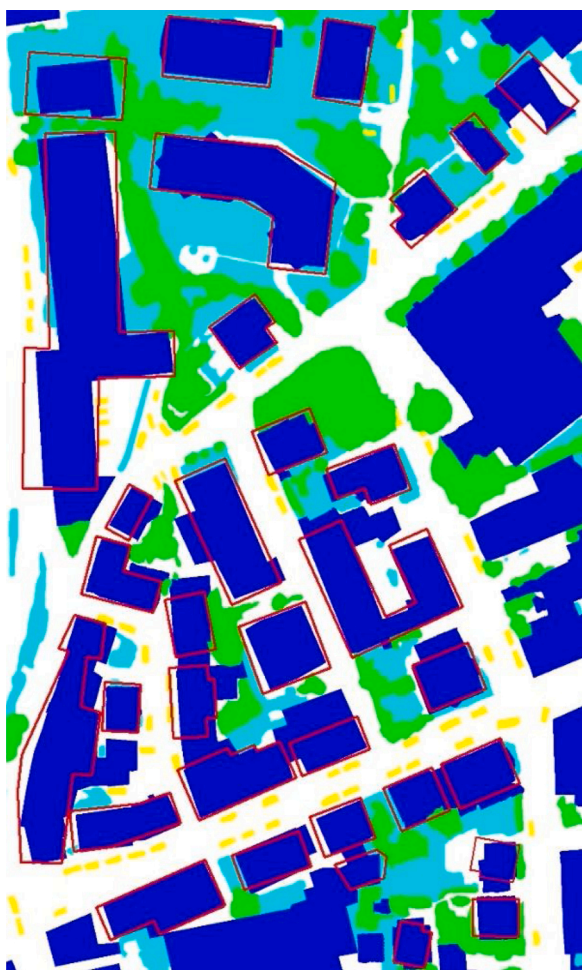


Fig. 10. Generated vector data of buildings superimposed onto the land cover map (reference). The categories of the land cover map are coded by colours: “building” (blue), “impervious surfaces” (white), “low vegetation” (cyan), “tree” (green), “car” (yellow). Source of land cover map: ISPRS WG III/4 (2014).

Mousa et al. (2019) where a combination of original aerial images and original laser scanning data have been used. This implies that the approach used is a comparable method for the automated mapping of buildings and other artificial topographical objects.

4.2. Software development

The programming of the applied method was done in “R” – a free software environment for statistical computing, data analysis, and graphics (R Core Team, 2019). The developed programs consist of several R-packages. For the DT classification task, the package “rpart” (recursive partitioning and regression trees) of the system library is used (Therneau and Atkinson, 2019). Here, the classification is expressed by a formula used in statistical modelling.

$$class \sim attribute1 + attribute2 + \dots \tag{5}$$

In this work, the dependent variable or response (*class*) is a name, and the independent variables (*attributes*) are numerical values. Both types of variables are vectors which contain the class names or the attribute values of all pixels.

The developed programs for enhancement of buildings allow an automated processing. The calculations are carried out in steps using small functions. Errors can then be discovered and eliminated more easily. Graphical display of the intermediate results can be used if

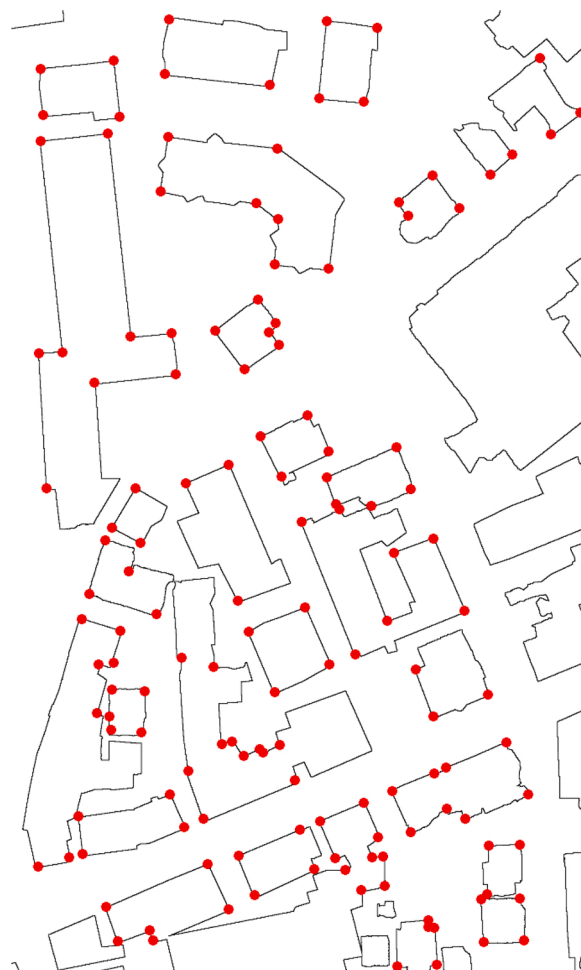


Fig. 11. Distribution of the check points along with the buildings generated from the provided land cover map (Vaihingen, Germany, area 1).

Table 3
Geometric accuracy of derived corner coordinates at Example 2.

reference	land cover map		ortho-image	
	x	y	x	y
coordinate	1.0	0.9	0.9	0.8
σ [m]	134		154	
n_{corner}	1		1	
$n_{gross\ error}$				

requested. Interactions by an operator are also possible. For example, small lines can be detected by measurement of a single pixel. Speed in processing has been gained by vector and matrix operations. Feature vectors with many attributes require computers with large main memories. Small computers may cause problems when applying AP profiles to multispectral images. The open-source image processing package, “EBImage”, is used for the enhancement of raster images (Andrzej et al., 2017). This R-package was originally developed for “microscopy-based cellular assays”, but the functionality provided can also be used for these tasks.

The developed programs regarding the enhancement of buildings are written in “R”. They comprise nine programs (‘extract all buildings’, ‘enhance image’, ‘extract single building’, ‘line detection’, ‘sequence of lines’, ‘adjustment of line’, ‘intersect corner points’, ‘adjustment of corner coordinates’, ‘plot results on references’). In addition, R-programs for the generation and assessment of LCM_2 developed in Höhle (2014) have been used for Example 2. The programs used for Example 1, generation of AP and LCM_1, are described in Dalla Mura et al. (2010)

Table 4

Geometric accuracy obtained by automated mapping of building outlines. d: distance between corresponding vertices, dp: perpendicular distance from vertex to boundary line, x,y: average of x- and y-coordinate error, NA: not announced.

author	year	source data	GSD [cm]	error definition	RMSE [m]	# of corners
Awrangjeb	2016	orig. ALS	100	d/NA	0.7	1013
Mousa et al.	2019	orig. images DSM (ALS)	8	dp/3 m	0.9	NA
Höhle	2020	ortho-images DSM, nDSM	9	x,y/ 3RMSE _{av}	1.0	270

and Damadoran et al. (2017), respectively.

5. Conclusions

The automated generation of vector maps has been accomplished by two different methods for the generation of land cover maps followed by one method for the extraction and enhancements of buildings. Both approaches could derive building vectors with a high cartographic quality. Their average geometric accuracy has been determined with $\sigma_{x,y,av} = 1.0$ m using numerous check points. The geometric accuracy can still be improved by more accurate input data (DSM, ortho-images, nDSM). Some editing will reduce the number of errors. The speed of operation must be improved by additional programming. The proposed methodology has the potential to automate the production and updating of topographic 2D maps. National and private mapping as well as crowd sourcing may benefit from such an automated approach.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRedit authorship contribution statement

Joachim Höhle: Conceptualization, Methodology, Software, Investigation, Visualization, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author would like to thank ISPRS Working Group III/4 (2012-2016) and prof. M. Gerke, TU Braunschweig, Germany, for providing test data. Thank is also given to prof. S. Lefèvre, Université de Bretagne Sud, Vannes, France (UBS), who introduced the author to attribute profiles, and to Dr. B.B. Damadoran, UBS, for implementing computer code for Example 1 of this article. Dr. M. Dalla Mura is thanked for supplying MatLab code for attribute profiles. Prof. M. Höhle, University of Stockholm, Sweden, supported the author in the design of R-software used for the generation of LCM and for the enhancement of buildings. The author thanks the editors and the anonymous reviewers for their valuable comments.

References

- Andrzej, O., Pau, G., Sklyar, O., Huber, W., 2017. Introduction to EBImage (accessed 12 June 2020). <https://www.bioconductor.org/packages/devel/bioc/vignettes/EBImage/inst/doc/EBImage-introduction.html>.
- Awrangjeb, M., 2016. Using point cloud data to identify, trace, and regularize the outlines of buildings. *International J. of Remote Sensing* 37 (January(3)), 551–579, 2016.
- Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. In: *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 23-4, pp. 110–117.
- Baltrusch, S., 2016. TrueDOP – a new quality step for official orthoimages. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLI-B4, pp. 619–624.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*. CRC Press.
- Congalton, R.G., Green, K., 2008. *Assessing the Accuracy of Remotely Sensed Data*, 2nd ed. CRC Press, Boca Raton, U.S.A. ISBN 978-1-4200-5512-2.
- Dalla Mura, M., Benediktsson, J.A., Waske, B., Bruzzone, L., 2010. Morphological attribute profiles for the analysis of very high-resolution images. *Ieee Trans. Geosci. Remote. Sens.* 48 (10), 3747–3761.
- Damadoran, B., Höhle, J., Lefèvre, S., 2017. Attribute profiles on derived features for urban land cover classification. *Photogrammetric Engineering & Remote Sensing*, vol. 83, no. 3, 183–193.
- Elberink, Oude, 2008. Problems in automated building reconstruction based on dense airborne laser scanning data. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII. Part B3a. Beijing 2008.
- Friedl, M.A., Brodley, C.E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sens. Environ.* 1997 (61), 399–409.
- Geoforum, 2011. *Spezifikation für ortofotos*, 3. edition, January 2011, p. 87 (accessed 12 June 2020). https://www.dropbox.com/sh/foct8tudoohnzy28/AABLAw7Yf5CD9DiQvFvzVE0Ha/Ortofoto-spezifikation?dl=0&preview=Geoforums+Ortofoto+spezifikation+2011+++3.udgave.pdf&subfolder_nav_tracking=1.
- Gerke, M., 2014. Normalized DSM. https://www.researchgate.net/publication/270104315_Normalized_DSM.
- Haala, N., Rothermel, M., 2012. Dense multi-stereo matching for high quality digital elevation models. *Photogrammetrie, Fernerkundung, Geoinformation* 2012 (4), 331–343.
- He, H., Zhou, J., Chen, M., Chen, T., Li, D., Cheng, P., 2019. Building extraction from UAV images jointly using 6D-SLIC and multiscale siamese convolutional networks. *Remote Sens. (Basel)* 2019 (11), 1040. <https://doi.org/10.3390/rs11091040>.
- Heipke, C., Woodsford, P.A., Gerke, M., 2008. Updating geospatial databases from images. In: Chen, Li, Baltsavias (Eds.), *Advances in Photogrammetry, Remote Sensing and Spatial Information Sciences: 2008 ISPRS Congress Book*. Taylor & Francis Group, London. ISBN 978-0-415-47805-2.
- Höhle, J., 2014. Generation of 2D land cover maps for urban areas using decision tree classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 15–21, 2014, II-7.
- Höhle, J., 2017. Generating topographic map data from classification results. *Remote Sens. (Basel)* 9, 224.
- ISPRS WG III/4, 2014. *2D Semantic Labelling Contest* (accessed 12 June 2020). <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html>.
- Kamusoko, C., 2019. *Remote Sensing Image Classification in R*. Springer Geography, ISSN 2194-315X. Springer Nature Singapore Pte Ltd., 2019.
- Konecny, G., Breitkopf, U., Radke, A., 2016. The status of topographic mapping in the world. A UNGGIM - ISPRS PROJECT 2012–2015. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 737–741. XLI-B4.
- Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M., Stilla, U., 2018. Classification with an edge: improving semantic image segmentation with boundary detection. *Isprs J. Photogramm. Remote. Sens.* 135, 158–172. January. 2018.
- Maxwell, A., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* 39 (9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>, 2018.
- Mayer, H., 1999. Automatic object extraction from aerial imagery-A survey focusing on buildings. *Comput. Vis. Image Underst.* 74 (2), 138–149.
- Mayr, W., 2011. UAV mapping—a user report. *ISPRS – Int. Arch. Photogramm. Remote Sens. Spatial Inform. Sci.*, XXXVIII-1/C22, pp. 277–282.
- Mousa, Y.A., Helmholtz, P., Belton, D., 2019. Building detection and regularisation using DSM and imagery information. *Photogramm. Rec.* 34 (March(165)), 85–107. <https://doi.org/10.1111/phor.12275>, 2019.
- R Core Team, 2019. *R: a Language and Environment for Statistical Computing*. R foundation for statistical computing, Vienna, Austria (accessed 12 June 2020). <https://www.R-project.org/>.
- Speckels, V., Syrek, L., Schlienkamp, A., 2010. DGPF-project: evaluation of digital photogrammetric camera systems – stereoplottung. *Photogrammetrie, Fernerkundung, Geoinformation* 2010 (2), 117–130.
- Tasar, O., Maggiori, E., Alliez, P., Tarabalka, Y., 2018. Polygonization of binary classification maps using mesh approximation with right angle regularity. *IGARSS 2018 – 2018 IEEE International Geoscience and Remote Sensing*, pp. 6404–6407.

Therneau, T., Atkinson, B., 2019. . Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-15. <https://cran.r-project.org/web/packages/rpart/rpart.pdf>.
Wang, Y., 2016. Automatic extraction of building outline from high resolution aerial imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLI-B3 419–423.

Wiki, O.S.M., 2020. OpenStreetMap Wiki (accessed 12 June 2020). <https://wiki.openstreetmap.org/wiki/>.