ROBUST BUILDING DETECTION IN AERIAL IMAGES

Sönke Müller*, Daniel Wilhelm Zaum

Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung, Universität Hannover, Germany -
{mueller, dzaum}@tnt.uni-hannover.de, *corresponding author

KEY WORDS: Building, Extraction, Aerial, Segmentation, Classification, Land Use.

ABSTRACT

The robust detection of buildings in aerial images is an important part of the automated interpretation of these data. Applications are e.g. quality control and automatic updating of GIS data, automatic land use analysis, measurement of sealed areas for public authority uses, etc. As additional data like laser scan data is expensive and often simply not available, the presented approach is based only on aerial images. It starts with a seeded region growing algorithm to segment the entire image. Then, photometric and geometric features are calculated for each region. Especially, robust form features increase the reliability of the approach. A numerical classification is performed to differentiate the classes building and non-building. The approach is applied to a test site and the classification results are compared with manually interpreted images.

1 INTRODUCTION

The detection of buildings is an important task for the interpretation of remote sensing data. Possible applications for automatic building detection are the creation and verification of maps and GIS data, automatic land use analysis, measurement of sealed areas for public authority uses, etc.

Buildings are significant objects in remote sensing data and directly indicate inhabited areas. In most cases, buildings are well recognizable by a human interpreter. An automatic system that is able to emulate a human operator is desired.

In the presented approach we concentrate on aerial images with a resolution of $0.3125\frac{\mathrm{m}}{\mathrm{pixel}}$, because often additional costs or simple unavailability prevent the utilization of additional sensor data.

The different approaches for building detection in remote sensing data differ in the used type of input data. Often, multi sensory data, e.g. SAR, infrared, stereo or laser scan images, is available as additional information that can improve the object extraction. Some approaches, like the presented one, work only on RGB images. The following section discusses some of them.

C. Lin and R. Nevatia (Lin and Nevatia, 1998) propose a building extraction method, that is based on the detection of edges in the image. It is assumed that the searched rectangular buildings can be distorted to parallelograms. The edges are taken as building hypothesis and classified by use of a feature vector and additional features like shadow. The found buildings are modeled as 3D objects.

G. Sohn and I. J. Downman (Sohn and Dowman, 2001) deal with building extraction in low-resolution images. They start with a Fourier transform to find dominant axes of groups of buildings, assuming that buildings are aligned parallel along a street. Due to the low image resolution, the building contours can only be found including gaps. Theses gaps are closed using the afore detected dominant axes. The regions found and relations between each other, are stored in a tree. The tree is used to find building structures.

The present approach is divided into a low-level and high-level image processing step. The low-level step includes image segmentation and postprocessing: first, the input RGB image is transformed to HSI and the intensity channel is taken as input for a region growing segmentation algorithm to get a segmented image. The seed points of this algorithm are set flexibly under consideration of the red channel. The segmentation result is postprocessed to compensate effects like holes in the regions and to merge roof regions which are separated into several parts. The regions are taken as building hypotheses in the following steps.

The high-level step includes feature extraction and final classification: first, a preselection is performed to reduce the number of building hypotheses by use of the region area and color. During the feature extraction, photometric, geometric and structural features are calculated for each hypothesis like:

- geometric features
  - object size: area, circumference
  - object form: roundness, compactness, lengthness, angles, etc.
- photometric features
  - most frequent and mean hue
- structural features
  - shadow, neighborhoods

Furthermore, the main axes of building hypotheses are calculated. They define a hexagon describing the region's contour. Eventually, a numerical classification is performed to decide whether a building hypothesis is a building or not.

The remaining part of this paper is organized as follows: in section 2 the initial segmentation procedure is described. Section 3 shows how and which features are used for the following classification step. The classification itself is described in section 4. The experimental results are presented in section 5 and the paper is summarized in section 6.

2 LOW-LEVEL PROCESSING

This section describes all low-level processing steps including preprocessing, where the input image is transformed, image segmentation, using a seeded region growing algorithm, and a postprocessing step, that allows merging of regions that fulfill special conditions.

2.1 Image preprocessing

The input images are available as raster data in RGB color space. To get a single band grey value image as input for the segmentation algorithm, the input RGB image is transformed to HSI. To get the intensity channel I of the HSI transformation, the following equation is used, where weights are set according to the perception of the human eye:

$$I = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \qquad (1)$$

The color angle H is calculated independent from the saturation and intensity and later used as a region feature:

$$H_1 = \arccos\left[\frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right] \qquad (2)$$

$$H = \begin{cases} 2\pi - H_1 & \text{, if } B > G \\ H_1 & \text{, else} \end{cases} \qquad (3)$$

The saturation S is calculated using equation 4:

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \qquad (4)$$

2.2 Seeded region growing algorithm

For the initial segmentation of the input image, a seeded region growing algorithm is used to find homogeneous roof regions in the image. The seed points are regularly distributed over the image with a seed point raster size set with respect to the expected roof size. For an input resolution of $0.3125\frac{m}{pixel}$, an appropriate raster size is 15 pixel to ensure that nearly every roof region is hit. This is not possible with a standard *split and merge algorithm*. As input channel of the region growing algorithm, the intensity channel is taken, calculated as described before in section 2.1. Attempts to use more than one channel as input for the region growing algorithm were made, but led to inferior results.

The seeded region growing algorithm starts at the pixel position of each seed point and compares this pixel's value with the neighboring pixel values. If the neighbor pixel values lie inside a given tolerance t_T, the neighboring pixels belong to the same region as the seed point. The region growing goes on recursively with the newly added pixels and ends, when no new neighboring pixels which fulfill the condition can be found. Pixels which already belong to a region are omitted. An example of an image segmentation made with the described procedure is shown in Fig. 2A.

To take only promising seed points, not a simple raster is taken as described up to now. The input image is divided into equally distributed regions of the size of the seed point raster. For each region a histogram of the red channel is calculated. Provided that building roofs are red, brownish or grey, the red channel indicates the existence of roofs.

For each region, an arbitrary pixel with a red channel corresponding to the maximum histogram value is chosen. In order to avoid seed points in shadow regions, this value has to be larger than a threshold t_S. Fig. 1 shows the red channel of a typical part of an input image and the corresponding histogram. The region growing step results in a completely segmented image.

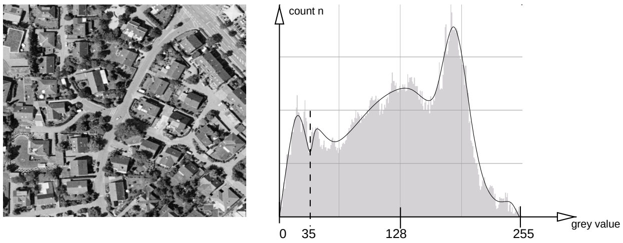

Figure 1: Red channel of a typical part of the input image, B) Histogram with threshold $t_S = 35$ for selection of seed points.

2.3 Image postprocessing

The aim of the postprocessing step is to improve the image segmentation for the following feature extraction and classification. The postprocessing consists of two steps. The first step is opening and closing of the segmented image. The second step is merging of special regions.

2.3.1 Opening and closing By use of opening and closing operators, the ragged edges of the regions (see Fig. 2) are smoothed and small holes in the regions are closed. As a consequence of the improvement of the region contours, the features used for classification of the regions can be calculated more precisely. In Fig. 2B, the result of opening and closing operations is shown.

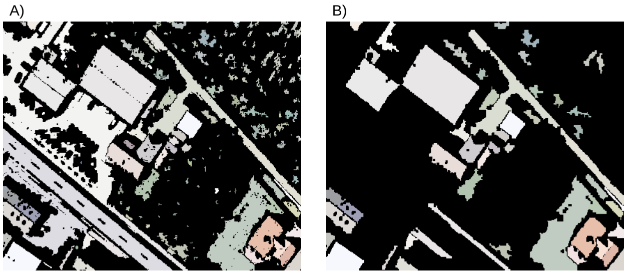

Figure 2: Example of the erosion and dilation step: A) image segmentation, B) image segmentation after applying of dilation and erosion, undersized and oversized regions are not considered.

2.3.2 Merging of regions The most important postprocessing step is the merging of regions belonging to the same roof. In Fig. 3 two examples of roofs that have been segmented into two different regions and the corresponding input images are given. Many roof regions are split at the roof ridge. Especially, complex roof structures like gable roofs, dormers, towers, roof-lights and superstructures lead to multiple regions for only one building, which complicates the subsequent classification.

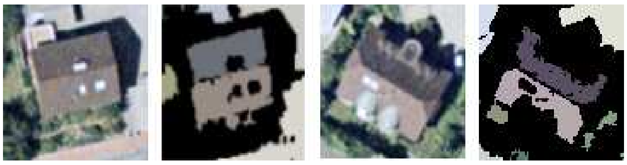

Figure 3: Two examples of roofs that consist of more than one region.

The complete merging procedure is depicted in Fig. 4. The leftmost image in Fig. 4 shows a symbolic example of an input image, that is the basis for the next steps.

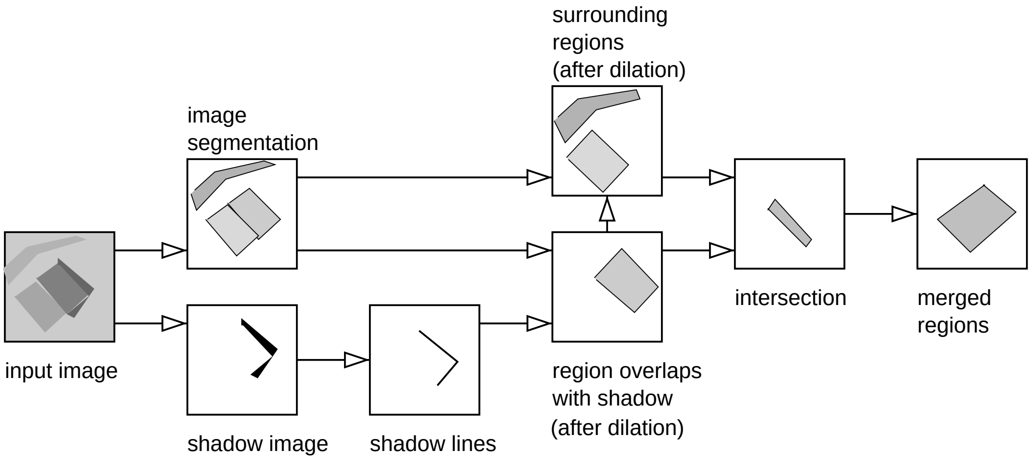
surrounding regions (after dilation)
image segmentation
input image
shadow image
shadow lines
region overlaps with shadow (after dilation)
intersection
merged regions

Figure 4: Steps of the merging procedure.

The segmented regions are dilated two times. Assuming that buildings cast a shadow, roof candidates with adjoined shadow regions are determined. Therefore, a shadow detection algorithm, performing a simple threshold operation as described in section 3.4.2, is applied to the input image. Subsequently, only shadow regions with straight borders are considered. The segmented regions are dilated two times. Regions now significantly overlapping with shadow regions, i.e. the intersection area is larger than a threshold S_1, are joined with overlapping neighbor roof candidate regions (see Fig. 4). Again, the intersection area has to be larger than a second threshold S_2. Eventually, the merged roof regions are eroded two times to restore the original region size.

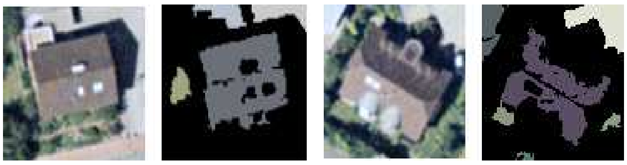

Figure 5: The result of a successful merging: labels of Fig. 3 merged together.

3 FEATURE EXTRACTION

This section describes the used numeric features which are extracted for each roof hypothesis. The result of the feature extraction, the numeric feature vector, is the basis for the following classification.

3.1 Preselection

To reduce calculation time, a preselection is performed, which sorts out implausible roof hypotheses. Roof hypotheses that are sorted out are eliminated and kept unconsidered during the classification. Therefore, the preselection has to be carried out carefully to prevent the loss of correct roofs. The features used for the preselection step are the region area and the mean hue angle.

3.1.1 Area The area of a roof hypothesis is calculated by counting the pixels of a region including holes. A hypothesis is assumed valid if the *area* fulfills equation 5.

$$30\,\text{pixels} < area < 25000\,\text{pixels} \tag{5}$$

3.1.2 Mean hue angle The second exclusion criterion is the color of a roof hypothesis. Therefore, the mean hue value of the pixels belonging to the corresponding region is calculated. Fig. 6 shows an example histogram of hue values for a roof region. The mean hue angle H_m is calculated as follows:

- in the range from $360°$ to $540°$ the hue value histogram is extended periodically
- if more than 5% of the pixels' hue values lie between $0°$ and $20°$ or $340°$ and $360°$ respectively:
  - calculate the mean in the range from $180°$ to $540°$
- else:
  - calculate the mean in the range from $0°$ to $360°$

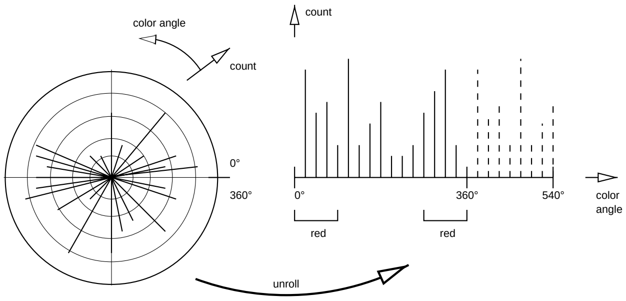

Figure 6: Example histogram of hue values for one roof hypothesis.

As explained before, roofs of buildings are perceptible in the red channel. All roof hypotheses with a mean hue value H_m close to green, see equation 6, are rejected.

$$115 < H_m < 125 \tag{6}$$

3.2 Geometric features

This section describes the geometric features used and how they are calculated.

3.2.1 Size features The size features of a roof hypothesis chosen are the *area* and *circumference*. The area is already calculated during the preselection step as described in section 3.1.1. The calculation of the circumference of a building is depicted in

Fig. 7: the left image shows a segmented roof candidate region, the right one shows its contour calculated with an outlining procedure. The counted contour pixel are used as circumference.

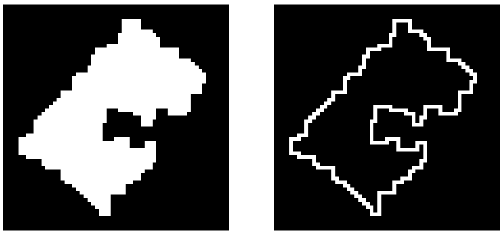

Figure 7: Outlining procedure: segmented roof region and region contour.

3.2.2 Form features

Form features are very important because the form can distinguish buildings from natural objects. Buildings are normally constructed with right angles. The features used are therefore *roundness*, *compactness*, *lengthness* and *characteristic angles* of a region.

The $roundness$ is calculated independently of the region's size as ratio of area to the square of the circumference. It ranges from 0 to 1.

$$roundness = \frac{4\pi \cdot area}{circumference^2} \tag{7}$$

The *compactness* of a region is defined as number of erosion steps that are necessary to remove the region in complete.

In (Baxes, 1994), different types of region axes are introduced. The main axis is defined as the line between two contour points that have the maximum distance among each other.

The two ancillary axes are defined as vertical lines to the main axis with the maximum distance from the contour to the main axis. For each side of the main axis one ancillary axis is defined.

The cross axis is defined as vertical line to the main axis that connects two contour points with the maximum distance to each other.

The axes calculation results in six points which lie on the contour of the investigated region. The corresponding hexagon approximates the region's shape. An example of such a hexagon is shown in Fig. 8.

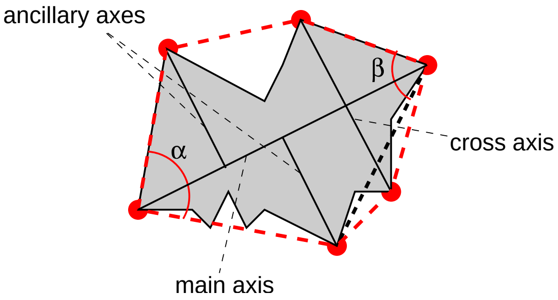

Figure 8: Approximating hexagon model.

The angles α and β, as depicted in in Fig. 8, are used as additional features. In most buildings these two angles are approximately right angles: on the left hand side of Fig. 9 a segmented building and its corresponding angles is illustrated. The described angles actually are approximately right angles. On the right hand side of Fig. 9 a segmented forest region is shown. Here, the angles are not close to being right angles.

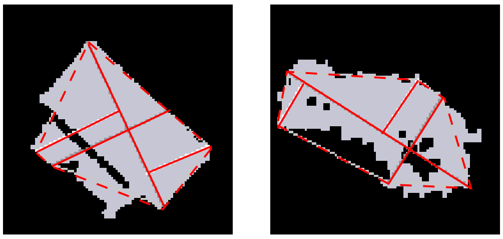

Figure 9: Examples for roof hypothesis.

The feature *lengthness* is also based on the calculated axes of a region. It is the ratio of the main axis length to the cross axis length.

Finally, it is measured how frayed a region is. Pixels that lie inside the hexagon and do belong to the investigated region (*hexagon area inside*) are counted, as well as the region pixels that lie outside the hexagon (*hexagon area outside*). The ratio of *hexagon area inside* to *hexagon area outside* is a feature that describes how frayed a region is.

Rectangular buildings with smooth contours fill the hexagon in complete, in contrast to e.g. parts of a segmented forest.

3.3 Photometric features

Two photometric features are calculated that are based on the hue angle histogram of the region pixels.

3.3.1 Hue angle

After applying a threshold on the intensity channel to discard shadow regions, the maximum value of the hue angle histogram of a region is taken as a feature.

3.3.2 Mean hue angle

Additionally, the mean hue angle of a building hypothesis as already described in section 3.1.2 is chosen as a feature.

3.4 Structural features

Structural features use information about neighboring regions and shadow.

3.4.1 Neighborhoods

Buildings are usually not freestanding but appear in groups. Consequently, buildings with other buildings nearby are more probable than single buildings. For each building hypothesis the neighboring buildings are counted. Neighbor means that the distance center to center is smaller than a threshold T_N. The number of neighboring buildings is used to support or reject unsteady hypothesis.

3.4.2 Shadow

The acquisition of aerial images requires good weather conditions, so the existence of shadow can be used for the extraction of buildings. Shadows are expected next to buildings, and therefore give hints to where buildings are.

The extracted shadow regions, shown in Fig. 10B, are obtained by a threshold decision on the intensity channel I of the input image. Only shadow regions with straight borders are considered. The first feature derived from extracted shadow is the general existence of shadow next to a building, the second is the ratio of region pixel that overlap with shadow, after a double dilation of the region.

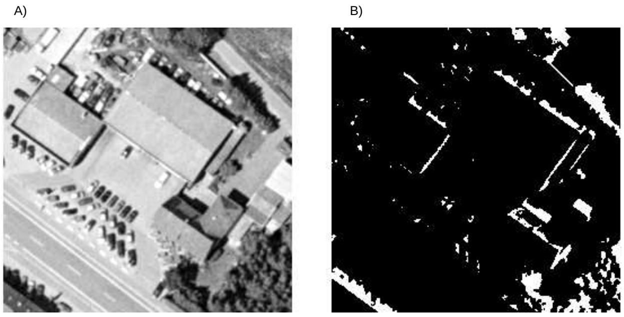

Figure 10: A) Input image in grey scale, B) Extracted shadow regions.

4 CLASSIFICATION

The classification is divided into a numerical and a contextual part, that are applied one after another. The context analysis is used as additional information to support the numerical classification. Result of the classification is a numeric rating in the range of $[0, 1]$.

4.1 Numerical classification

The numerical classification is carried out first. All numeric features, described in section 3 are taken as input vector for a linear regression classifier. A more detailed description of the used classifier and a calculation scheme can be found in (Meyer-Brötz and Schürmann, 1970).

Table 1 gives an overview of the used feature vector. The learning sample was generated by manual classification of a test dataset and consists of about 120 representative feature vectors for each class, building and non-building. The a-priori probabilities for the two discriminated classes are set to 0.5.

feature	occurring values for buildings in learning sample
region adjoins shadow	$0 < x$
mean hue angle	$0 < x < 43$ or $190 < x < 360$
hue angle	$0 < x < 43$ or $180 < x < 360$
region area	$200 < x < 5000$
area/circumference	$2 < x < 10$
lengthness	$3 < x < 4$
compactness	$3 < x < 15$
roundness	$x < 0.3$
area/ shadow area	$x < 0.8$
hexagon area inside/outside	$0.8 < x < 1.1$
form angles	$88 < x < 93$

Table 1: List of numeric features used.

Only regions that have passed the preselection described in section 3.1 are used as building hypothesis and classified.

4.2 Contextual Classification

The numerical classification results in a probability for the class building. This is used as input for the contextual classification step. All building hypotheses having a probability below a threshold are assigned to the class non-building. For the others the surrounding building hypotheses are counted. If no neighboring buildings based on the numerical classification and the threshold decision exist, the probability for the investigated building hypothesis itself is reduced by 0.1. If at least one building hypothesis region is in proximity, the probability for the investigated building hypothesis itself keeps the value of the numerical classification.

5 RESULTS

This section shows and discusses the results of the proposed approach. Additionally, the validity of the approach is discussed.

5.1 Evaluation Method

The evaluation of the approach is based on manually segmented buildings in three test images ($\approx$ 53ha), cp. Fig. 11.

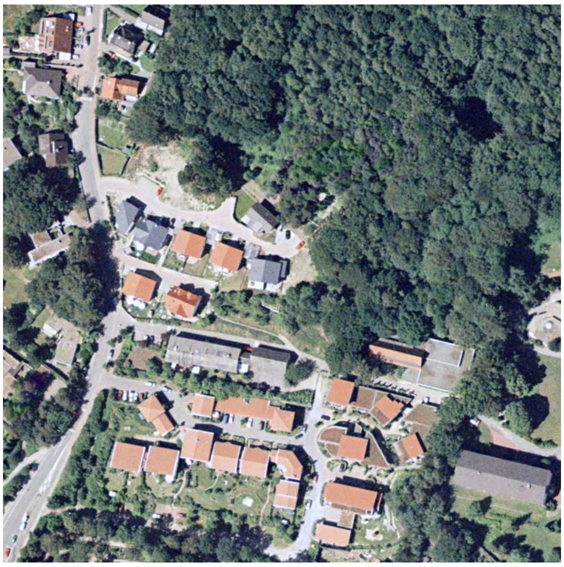

Figure 11: Part of an original image, resolution $0.3125 \frac{\mathrm{m}}{\mathrm{pixel}}$.

Two measurements for a detection evaluation as described in (Lin and Nevatia, 1998) were made:

$$detection\ \ percentage = \frac{100 \cdot TP}{TP + TN} \tag{8}$$

$$branch\ \ factor = \frac{100 \cdot FP}{TP + FP} \tag{9}$$

Two measurements are calculated by comparing the manually detected buildings and the automatic results (cp. Fig. 12), where TP (true positive) is a building detected by both a person and the automatic approach, FP (false positive) is a building detected by the automatic approach but not a person, and TN (true negative) is a building detected by a person but not by the automatic approach. A building is rated as detected, if at least a small part of it is detected by the automatic approach; alternatively, it could be required that at least a certain fraction of the building area has to be detected.

The $detection\,percentage$ (DP) describes how many of the existing buildings in the scene are found by the automatic approach, the $branch\,factor$ (BF) gives a hint on how many buildings are found erroneously. The DP is 100% if the whole image is classified as class building. In this case also the BF would be very large. The goal is to maximize the DP while keeping the BF low.

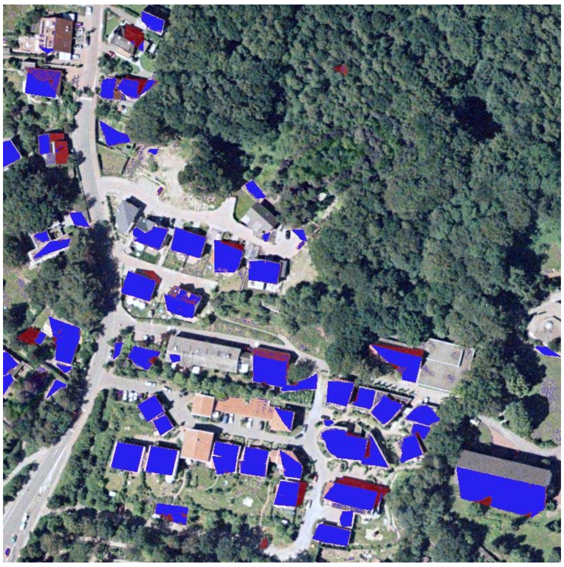

Figure 12: Classification result: Building polygons marked in blue.

5.2 Evaluation results of a test dataset

The analysis is tested on a set of three images of about 1500×1200 pixels. Each image contains about 120 buildings. The results are shown in detail in Table 2 and the mean values for the whole test site in Table 3.

image	TP	FP	TN	DP	BF
W_1	89	28	18	83.2%	23.9%
W_2	109	34	24	82.0%	23.8%
W_3	85	21	31	73.3%	19.8%

Table 2: Evaluation results of three test images.

mean values of images	DP	BF
W_1, W_2, W_3	79.5%	22.5%

Table 3: Mean results of three test images.

5.3 Validity of the results

The classification uses a knowledge base that was optimized to the test data set of a rural area. The approach runs satisfactorily in regions of small towns with characteristic one family houses, small apartment houses, and industrial areas. One aspect of the classification was to reliably detect or reject vegetation areas, that are not dominant in downtown areas. The knowledge base is also not optimized for detection of multistory buildings and town houses. Due to the modular structure of the proposed approach, the knowledge base can be easily expanded to other situations.

The approach was additionally tested on a set of IKONOS images with a spatial resolution of 1.0m. This test was done without manual parameter tuning. Due to the lack of a precise manual segmentation, a numerical evaluation was not done, but the results look comparably good to those of the tested aerial images (see Fig. 13). Oversegmentation caused by not optimally adapted thresholds during low-level processing does not affect the detection result, since it is based on robust features.

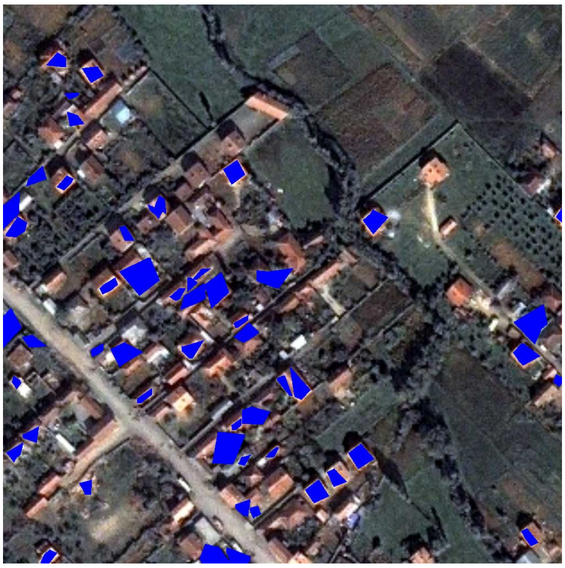

Figure 13: Classification result of an IKONS image: Building polygons marked in blue.

6 CONCLUSIONS

We propose an algorithm to detect buildings in aerial images. A seeded-region growing algorithm is used to segment the entire image. A preselection reduces the feature extraction to only plausible hypothesis. The classification is based on robust features, especially form features. The numerical linear regression classifier is extended by a contextual classification, considering the surroundings of a building.

To evaluate the building detection approach, it is tested on a site of ≈ 53ha. The results in Table 3 show that the proposed approach is applicable. Other approaches sometimes achieve smaller $branch\,factors$. However they mostly concentrate on small images. This leads to a smaller $branch\ \ factor$. In the present approach, test sites including large vegetation areas are used. In consideration of today's available computing power, representative test sites have to be tested to get expressive results.

The runtime of the present approach depends on the image content. The low-level processing requires the major part of the program runtime. The algorithm's runtime on a 2.8GHz Pentium4 computer is 45 to 75 minutes for an image of 6400 x 6400 pixels.

REFERENCES

Baxes, G. A., 1994. Digital image processing: principles and applications. John Wiley & Sons, Inc., New York, NY, USA.

Lin, C. and Nevatia, R., 1998. Building detection and description from a single intensity image. Computer Vision and Image Understanding: CVIU 72(2), pp. 101–121.

Meyer-Brötz, G. and Schürmann, J., 1970. Methoden der automatischen Zeichenerkennung. K. Becker-Berke and R. Herschel, R. Oldenbourg Verlag, München-Wien.

Sohn, G. and Dowman, I. J., 2001. Extraction of buildings from high resolution satellite data. In: Proc. ASCONA 2001, Ascona, Swiss, pp. 345–355.