COMPARISON OF METHODS FOR AUTOMATED BUILDING EXTRACTION FROM HIGH RESOLUTION IMAGE DATA

G. Vozikis

GEOMET Ltd., Faneromenis 4, 15561 Holargos-Athens, GREECE
george.vozikis@geomet.gr

KEY WORDS: Photogrammetry, Building , Detection , Transformation, Model, Pattern

ABSTRACT:

This paper discusses a comparison analysis of different methods for automated building extraction from aerial and spaceborne imagery. Particularly approaches employing the Hough Transformation, Pattern Recognition Procedures and Texture Analysis are examined. Throughout this investigation advantages and disadvantages of the mentioned methods are examined, in order to see which procedures are suitable for extracting the geometric building properties, and thus to automatically create a DCM (Digital City Model). The examined data sets consist of panchromatic imagery coming from both very high resolution satellites, as well as line scanning aerial sensors. A quantitative and qualitative assessment will help to evaluate the previously mentioned procedures.

1. INTRODUCTION

Automated building extraction from high resolution image data (either airborne or spaceborne) is becoming more and more mature. Everyday new techniques are investigated and the results are getting more and more reliable, while the degree of automation increases. Each building extraction method is of course coupled to certain pros and cons. The use of the Hough Transformation has proven to be a very promising tool in the frame of the automated creation of Digital City Models (DCMs), by extracting building properties from optical data. But also approaches based on Image Matching or Texture Analysis seem to provide usable results. A DCM is described through the outlines of buildings outlines of an urban area. Vertical walls are assumed, and the elevation information of these buildings can be taken from a DSM (Digital Surface Model). The creation of the DSM and the assignement of the elevation value is not discussed in this paper, thus when mentioning DCMs we actually mean the Model that holds the 2D outline-information of a building.

The goal of this paper is to conclude for which kind of data sets and accuracy pretensions a certain approach is recommendable. Moreover, the reachable degree of automation is also examined, in order to see how reliable results are that were produced without human interaction.

Table 1: Examined data sets.

Sensor	Location	GSD (m)	Extents (km)
ADS40	Valladolid, Spain	0.25	1 x 1
HRSC-AX	Bern, Switzerland	ca. 0.3	0.2 x 0.3
Quickbird	Denver, USA	0.6	16.9 x 16.5
IKONOS	Athens, Greece	1	9.7 x 12.3
Orbview 3	Orange, USA	1	0.6 x 0.7

Altogether, five different datasets, coming from airborne and spaceborne sensors, were examined. These datasets depict urban regions with varying building sizes, patterns and densities. It should be mentioned here that only subsets have been used for the investigations.

2. DESCRIPTION OF WORKFLOWS

2.1 Hough Transformation

The proposed workflow for automated building extraction from image data by employing the Hough Transformation has been thoroughly described in Vozikis (2004). Figure 1 shows the major steps of the process.
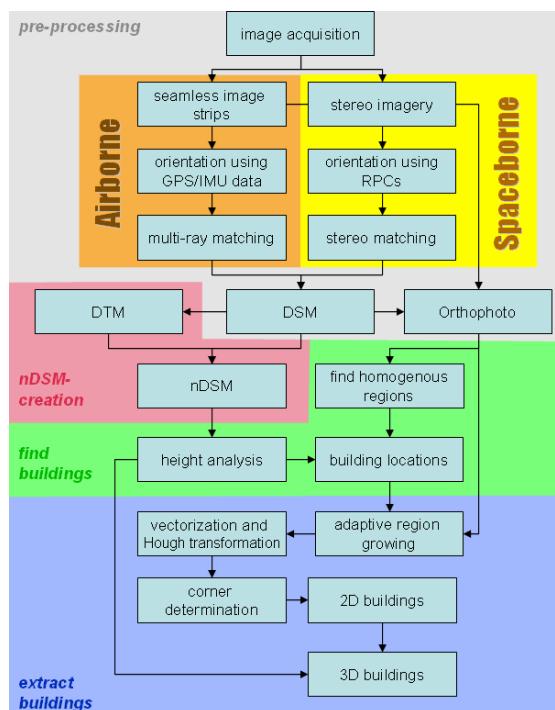

Figure 1: Proposed workflow for automated building extraction from image data

All steps in this workflow are highly automated and human interaction is reduced to a minimum.
In the following the 4 major steps are briefly described.

2.1.1   Pre-Processing

This step comprises the procedures from orientation of the input data up to the DSM (Digital Surface Model) creation. For VHR satellite imagery the orientation approach is based on the RFM (Rational Function Model) (Vozikis et al., 2003). When dealing with aerial imagery it is made use of GPS/INS information in order perform direct georeferencing, and thus automated image triangulation (Scholten and Gwinner, 2003). The DSM extraction is performed by automated correlation procedures, which nowadays are very mature and produce very good results.

2.1.2   nDSM Creation

The goal is to derive the DTM (Digital Terrain Model) from the DSM and subtract it from the DSM in order to produce the so-called nDSM (normalized Digital Surface Model). This way all extruding objects in the data set (including buildings) stand on elevation height 0 (Figure 2).
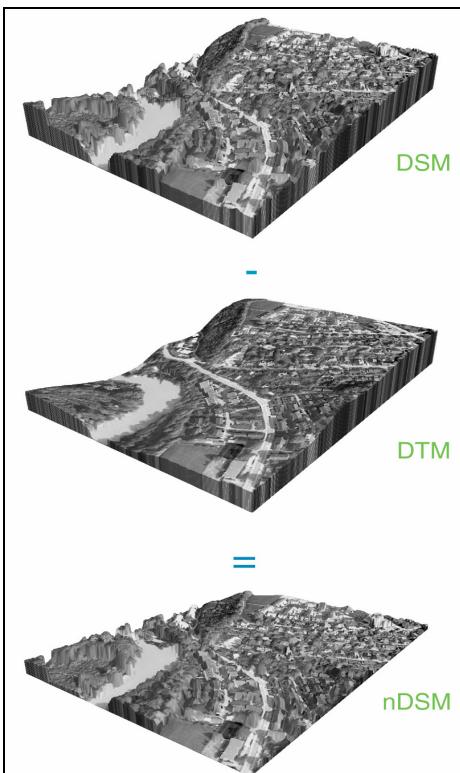

Figure 2: DSM, DTM and nDSM.

2.1.3   Building Detection (Seeding)

This crucial step deals with the identification of potential building candidates in the data sets (=determination of seed points inside buildings). It is proposed to perform 2 statistical analyses. First, perform a thresholding in the nDSM and filter out all objects that are not taller than a certain height, and second, perform texture analysis in the image data to keep only roof-similar regions in the data set (Vozikis, 2004).
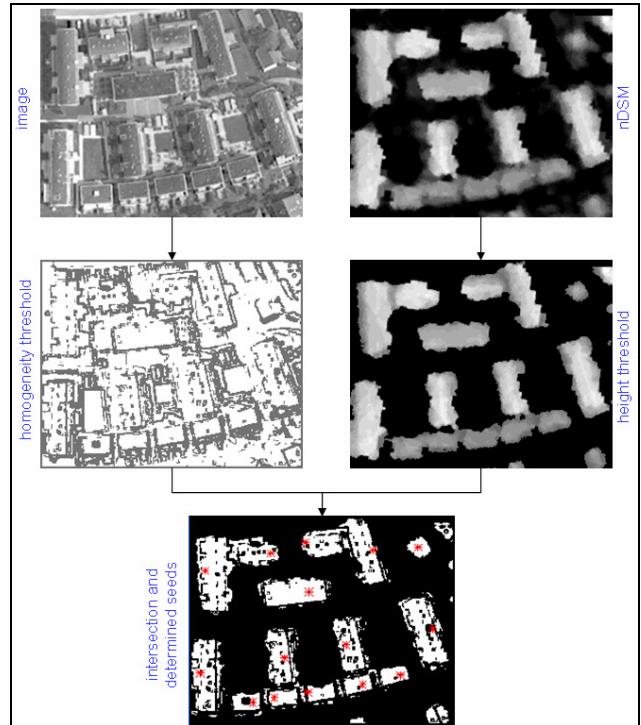

Figure 3: Computation of seed points (red asterisks) inside potential building candidates by height-thresholding and texture filtering.

2.1.4   Building Extraction

By applying the Hough Transformation (to an image of gradient or of contours) the geometric properties of the buildings (building edges and corners) are extracted. Our approach is based on a stepwise, iterative Hough Transformation in combination with an adaptive region growing algorithm (Vozikis and Jansa, 2008). The general idea is to transform the information in the image (feature space) into a parameter space and apply there an analysis. It is a technique for isolating features that share common characteristics. The classical Hough transformation is used to detect lines, circles, ellipses etc., whereas the generalized form can be used to detect features that cannot easily be described in an analytical way.

The mathematical analysis of the Hough Transformation is described in detail in Gonzalez and Woods (1992).

Briefly it can be described as follows:

$$\rho = x\cos(\theta) - y\sin(\theta) \tag{1}$$

where ρ is the perpendicular distance of a line from the origin and θ the angle (in the range 0 to π) as illustrated in Figure 4.

To apply this function on the whole image, Equation 1 can be extended as shown in Equation 2.

$$H(\theta,\rho) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} F(x,y)\delta(\rho - x\cos(\theta) - y\sin(\theta))dxdy \tag{2}$$

where δ is the Dirac delta-function. Each point (x,y) in the original image F(x,y) is transformed into a sinusoid $\rho = x\cos(\theta) - y\sin(\theta)$.
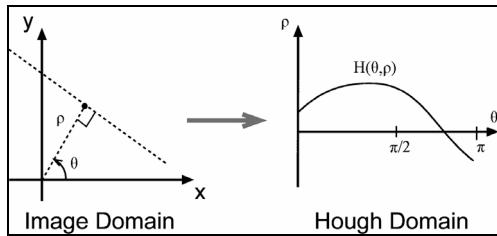
Figure 4: Hough transformation.

Points that lie on the same line in the image (feature space = Image Domain) will produce sinusoids that all intersect at a single point in the Hough domain (parameter space = Hough Domain). For the inverse transform, or back-projection, each intersection point in the Hough domain is transformed into a straight line in the image (Figure 5).
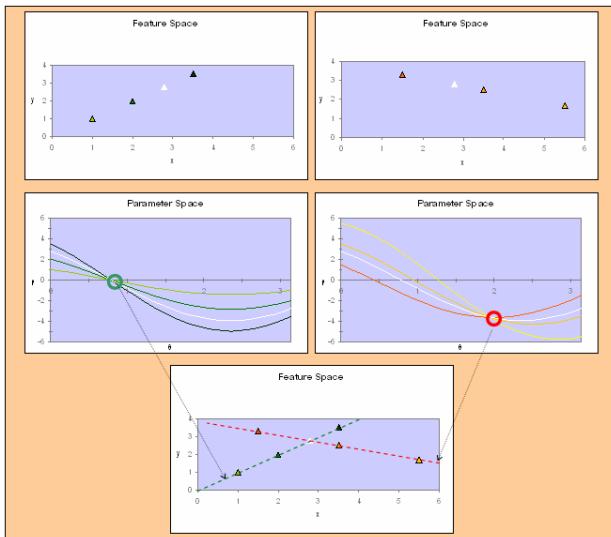

Figure 5: Example: Hough transformation.

The research shows that by using the Hough Transformation for building extraction we have many advantages, such as the good handling of noisy data, the easy adjustment of level of detail of the output data, the ability to force certain geometric properties into the extracted buildings and the possibility to bridge gaps, meaning that building corners that might not be visible in the imagery can be determined accurately. The proposed methodology proves to have certain weaknesses when dealing with radiometrically heterogeneous roofs, when big shadows cover large areas of roofs of the buildings to be extracted, when the building geometry becomes very complex, or when the input data set comprises many compound building (Vozikis and Jansa, 2008).

2.2 Image Matching

This strategy follows the basic principle of image matching by correlation. A given reference image matrix is searched in the image under investigation (the so-called search image) by moving the reference matrix pixel by pixel over the entire image area. Potential candidate positions, i.e. positions of high similarity, are marked if a so-called correlation coefficient exceeds a predefined threshold. In order to find the optimum geometric fit, the searching procedure includes, besides translation, also rotation and scaling. Thus houses of similar shape but different size are found too.

The reference image is usually a small image matrix, here depending on the size of the building to be searched, whereas the search image is a rather big image matrix in our case covering the whole area under investigation.
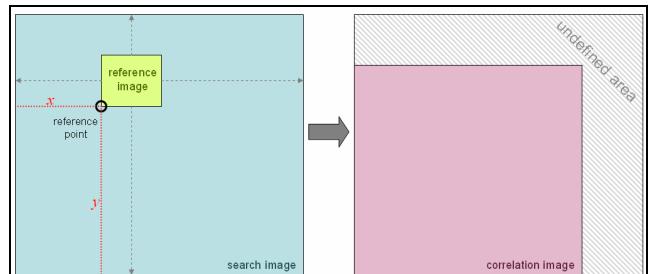

Figure 6: Search, reference and correlation image.

Figure 6 shows the principle of the correlation procedure. The left hand side indicates the searching process with the reference image and the given spaceborne or airborne image as search image. The correlation index is computed for each position of the reference image and the results are stored as similarity measure in the so-called correlation image. Potential building positions are characterized by a high correlation coefficient and thus the correlation image just needs to be thresholded and the local maxima are localised. It has to be mentioned that one crucial parameter is certainly the appropriate threshold value. Its choice determines quite significantly the quality of the result. If the threshold is too low, too many buildings are detected leading to a great number of false matches. If the threshold is too high, the selection is too strict and, as a consequence, too many buildings will be rejected. It is not possible to define an optimum threshold as a general suggestion. For the cross-correlation coefficient using 0.7 to 0.8 is certainly a good choice for starting, but individual adjustments are necessary in any case.

As measure of similarity the cross-correlation coefficient (Equation 3) is adopted, but also other measures can be used (Equations 4 and 5).

$$c_1 = \frac{\sum (g_1 - \bar{g}_1) \cdot (g_2 - \bar{g}_2)}{\sqrt{\sum (g_1 - \bar{g}_1) \cdot \sum (g_2 - \bar{g}_2)}} \qquad (3)$$

$$c_2 = \frac{\sum (|g_1 - g_2| / (g_1 + g_2))}{n} \qquad (4)$$

$$c_3 = \sqrt{\frac{\sum (g_1 - g_2)^2}{n}} \qquad (5)$$

where $g1$ and $g2$ are the grey values in the reference and search window,
$\bar{g}_1$ and $\bar{g}_2$ are the mean grey values in the reference and the search window and
n is the number of used pixels.

Kraus (1996) suggests rewriting Equation 3 as follows for a more efficient computation:

$$c_1 = \frac{\sum g_1 \cdot g_2 - n \cdot \bar{g}_1 \cdot \bar{g}_2}{\sqrt{\sum (g_1^2 - n \cdot \bar{g}_1^2) \cdot \sum (g_2^2 - n \cdot \bar{g}_2^2)}} \qquad (6)$$

Note, that when using Equation 6 the computing effort is reduced since the expression $\sum (g_1^2 - n \cdot \bar{g}_1^2)$ is constant during the whole process and has to be calculated only once.

The finding of the maxima in the correlation image with sub-pixel accuracy by approximating the discrete correlation function by a continuous polynomial function is broadly discussed in Kraus (1996) and will not be described here.

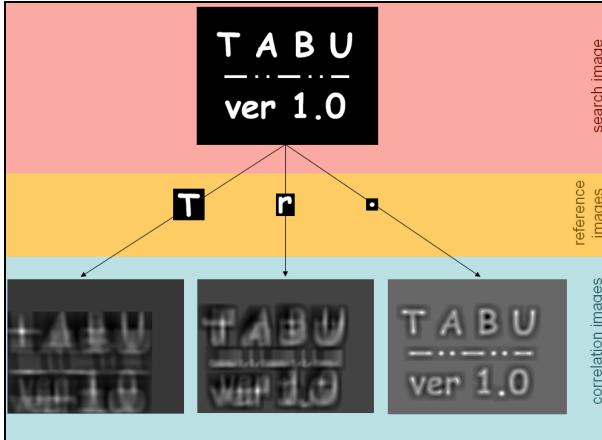Figure 7 shows an example were one search image was matched with multiple reference images.

Figure 7: Correlation example. The brighter the pixels, the higher the correlation, i.e. the greater the similarity between search image and reference image.

For the practical implementation, the reference images of buildings are stored in a library. For each of these buildings also vector information (describing the building outline in the reference image coordinate system) is available. Thus the library contains multiple building types to which one image patch and one vector representation corresponds.

Once a location of high correlation is found in the search image, the vector data of this building is transformed into the coordinate system of the search image.
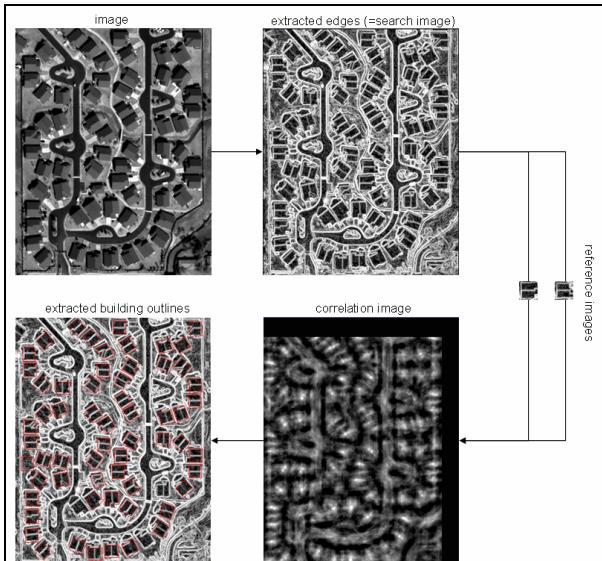

Figure 8: Example on image matching with two reference images and multiple orientations (Quickbird-subset of Phoenix area). Note that the computed correlation image is a multi-dimensional image. The number of dimension corresponds to the number of rotations (here: 120 with a rotation step of 3 degrees).

It is often the case that buildings of the same (or similar) shape have different colours (grey values) in the images (e.g. due to different roof materials). Therefore it is advisable not to store image patches of the investigated buildings in the library, but instead, register their edges. In this case, also the search image has to be edge-extracted before applying the matching procedure. For gaining the edge information, classical operators like the Canny edge detector, Sobel operator, Laplacian of Gaussian etc. can be applied.

Figure 8 shows an example of the image matching procedure.

2.3 Texture Analysis

One of the simplest ways for describing texture is to use statistical moments of grey level histograms of an image or a region. Measures of texture computed using only histograms suffer from the limitation that they carry no information regarding the relative position of the pixels with respect to each other. One way to bring this type of information into texture analysis process is to consider not only the distribution of intensities, but also the distribution of intensity variation (Gonzalez and Woods, 2002).

For this kind of textural examination, firstly the so-called co-occurrence matrix has to be derived for the examined area. This particular matrix holds e.g. information of pixel changes in multiple directions (usually horizontally, vertically and diagonally). The co-occurrence matrix' extents are same in both directions and equal to the number of grey levels that will be considered. For example, for an 8 bit image (256 grey values) the co-occurrence matrix' extents would be 256 by 256. Usually a recoding is carried out to reduce the number of grey value classes (also called bins). A recoding of the original image down to 16 grey levels is for most of the cases satisfying (Gong et al., 1992). Nevertheless, during this research (on texture analysis), all images were recoded to 40 bins.

Figure 9: Image and corresponding co-occurrence matrices in horizontal (left to right) and vertical (top to down) directions.

At each position mc,r the co-occurrence matrix holds the number of changes from class r (=row indices) to class c (=column indices) (see Figure 9). This computation is carried out for multiple directions, meaning that one co-occurrence matrix is created for each direction. Figure 10 illustrates the creation of such matrices; here, four grey values exist and the co-occurrence matrices are derived for horizontal and vertical directions.

The task now is to analyze a given co-occurrence matrix in order to categorize the region for which it was computed. Therefore descriptors are needed that characterize these matrices. Some of the most commonly used descriptors are

thoroughly described in (Haralick 1979, Gonzalez and Woods 2002, Zhang 2001). It is obvious how important it is to include some kind of information that tells us whether the values are well distributed over the whole matrix, or whether they are mostly located close to the matrix diagonals (e.g. Difference Moment or Inverse Difference Moment, Equations 5-10 and 5-11). In case big values lie close or on the diagonal (Figure 10, non-urban), the region under investigation is expected to be homogeneous, whereas if the values are distributed more homogeneously (Figure 10, urban), the co-occurrence matrix corresponds to a heterogeneous region.
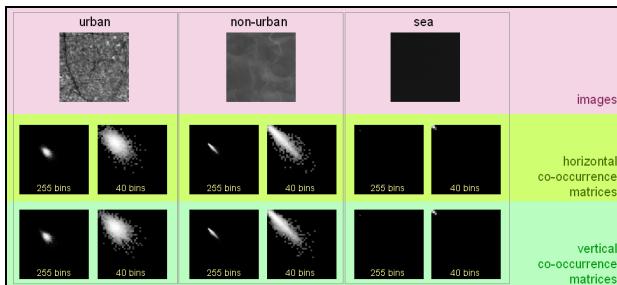

Figure 10: Horizontal and vertical co-occurrence matrices for 3 different types of terrain.

Usually the quality of the input data (i.e. imagery and DSM) is responsible for erroneous results. It is very difficult for the introduced algorithm to produce correct results, if the buildings to be extracted does not cover a certain number of pixels. For instance, when dealing with IKONOS imagery (with GSD of 1m) a small house of 8m x 10m will most probably not be extracted correctly.

The procedures discussed in paragraph 2.2 and 2.3 fall into the category of feature extraction and make use of the imagery for deriving the geometric building properties.
The general idea is to use a library where characteristic features of buildings are stored; by analyzing the image, areas are searched that correspond to a high degree to the registered "library buildings". Characteristic features of a building can be textural measures (by using the so-called occurrence and co-occurrence descriptors) or similarity measures of the image grey values.

3. RESULTS

In this chapter an evaluation of the presented methods is given. It consists of a quantitative and qualitative description, and moreover shortcomings and weaknesses of the presented methods are discussed.

Since many subsets are examined that are coming from various types of line scanning systems, both airborne and spaceborne (ADS40, HRSC-AX, Quickbird, Orbview, IKONOS, SPOT5), their outcomes will not be listed individually. Errors in the qualitative evaluation will be given in image space units (pixels).

We consider that the pre-processing has been carried out without error, so that the orientation of the imagery and the derived orthophotos on which we apply the investigated techniques are correct. We will also not evaluate nDSM extraction algorithms and their qualities in detail, since this is not topic of this research.

The presented outcomes are divided into two groups: quantitative and qualitative results. Moreover, the three presented DCM extraction approaches are evaluated individually. Input data is subdivided into categories depending on image scale and building density (low urban and urban) of the investigated areas. Image scale is defined as the scale that we would expect from an analogue product, e.g. for a 1:10,000 product we expect 1-2 metres accuracy in nature, if the graphical accuracy and visual perceptivity are 0.1-0.2mm.

Regarding the mentioned image scales the three interpretation categories are:
1. scale A: 1:1000-1:4000
2. scale B: 1:4000-1:12000
3. scale C: < 1:12000

3.1 Quantitative Assessment of Building Extraction

The aim in the quantitative analysis is to evaluate whether the presented approaches are practical in sense of completeness of building detection of the result, i.e. how many buildings were actually found. It is investigated whether the techniques for finding potential building candidates are applicable. Furthermore, an evaluation is carried out to see how many of these buildings were extracted and to what a degree:

- **CFB**: **C**orrectly **F**ound **B**uildings,
- **NFB**: **N**ot **F**ound **B**uildings (also includes insufficiently mapped buildings: building seed point was determined successfully, but the adaptive region growing process did not manage to create an area that covers a reasonable amount of the object),
- **WFB**: **W**rongly **F**ound **B**uildings, i.e. found objects were in reality no building exists.

The calculation of CFB (true positive), NFB (false negative) and WFB (false positive) are briefly explained in the following: The CFB and NFB percentages are calculated with respect to the total number of existing buildings in the area under investigation, whereas the WFB is calculated with respect to the total number of found buildings (comprising correctly and wrongly found buildings). Figure 11 shows the way of computing and a numerical example, respectively.

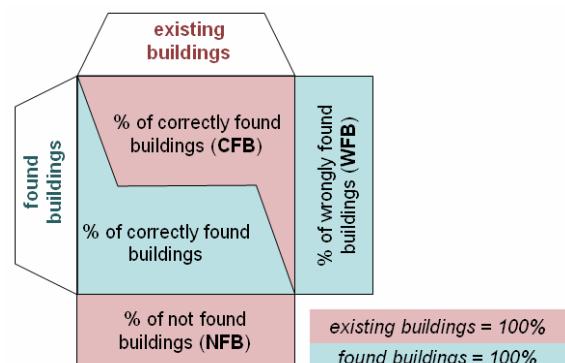Figure 11: Illustration for quantitative assessment computation.

For the evaluation of the outcomes altogether 13 different scenes containing 677 buildings were examined.
Table 2 allocates the quantitative analysis.

Concerning the level of detail that can be derived from individual data sets the Nyquist theorem has to be taken into consideration ("Sampling rate must be at least twice as high as

the highest frequency of the signal."). For our situation it means that sensor's geometric resolution determines the object resolution, or in other words, the level of detail of the object.

Table2: Quantitative results.

scale	A	B	B	C	C
type	urban	low urban	low urban	low urban	low urban
sensors	airborne	airborne	space-borne	space-borne	space-borne
method	Seeding	Seeding	Seeding	Matching	Texture Analysis
CFB	80.5%	90.2%	97%	88.3%	34.4%
NFB	19.5%	9.8%	3%	11.7%	65.6%
WFB	7.3%	4.8%	30.3%	1.7%	17.2%

3.2 Qualitative Assessment of Building Extraction

The qualitative analysis is based on a comparison between the building outlines derived by using the proposed automated methodology, and manually mapped buildings (image restitution). The manual mapping is carried out by a professional operator who performs 2D (or 3D) digitization on the input data (oriented imagery or orthophotos).

The residuals of each building corner from the manual mapping and the closest point of the automatically extracted shape are computed as a quality measure.

The results are categorized in three groups depending on the method used for building extraction (Table 3). The RMS is given in pixels.

Table3: Qualitative results.

	Hough	Matching	Texture Analysis
RMS x	0.937	0.898	0.954
RMS y	0.914	0.958	0.996
total RMS	1.309	1.313	1.379

The number of examined objects is the same as in the quantitative analysis.

Note that the figures in Table 3 are based on image residuals. They show the difference of the automatically derived corner points and the digitized ones in the image. As our data sets were acquired with vertical viewing angles these results can be also interpreted as planimetric object space residuals.

But when dealing with images that were captured with oblique viewing angles, the buildings must be projected into object space in order to carry out a qualitative analysis in the reference system.

4. CONCLUSIONS

The aim of this work was to propose a method for generating DCMs which makes use of images from spaceborne or airborne line scanning devices, on orthophotos if available and on elevation models. Various image processing techniques, such as Hough transformation, adaptive region growing, image matching, texture analysis, were employed and investigated for deriving the strengths and weaknesses of each. A variety of data sets were tested, coming from both spaceborne and airborne acquisition systems. Through the research based on adaptive region growing and on the iterative Hough transformation we can conclude that the method is very powerful, but has also some weaknesses. One is the high dependence on the radiometric quality of the input imagery. Furthermore, rather small buildings will not be treated correctly. Image matching proved to be a very effective, but very time consuming. The

suggested strategy of texture analysis, although very efficient for pattern recognition over areas in small scale imagery, was not very successful for extracting individual buildings.

Through this research partly very good results were obtained, but nevertheless further investigations are necessary for improving the quality of the results even more.

Future work will be focused on:

- Extraction of objects with holes (e.g. houses with inner courtyards), i.e. deriving the inner and outer boundary of buildings.
- Research on constraint settings for aggregating neighbouring roof parts that belong to one building.
- Introduction of multispectral information for making the algorithms more efficient, especially as far as seed point determination is concerned.
- Extract edges on sub-pixel bases.
- Integrate a hierarchical approach in order to decrease computation time.

5. REFERENCES

Gong, P., Marceau, D.J. and Howarth, P.J., 1992. A Comparison of Spatial Feature Extraction Algorithms for Land-Use Classification with SPOT HRV Data. Remote Sensing of Environment, Vol. 40, pp. 137-151.

Gonzalez, R.C., and Woods, R.E., 1992. *Digital Image Processing*. Reading, MA: Addison Wesley.

Gonzalez, R.C. and Woods, R.E., 2002. Digital Image Processing, Second Edition. Patience Hall, New Jersey, 799 pages.

Haralick, R.M., 1979. Statistical and Structural Approaches to Texture. Proceedings IEEE, Vol. 67, No.5, pp. 786-804.

Kraus, K., 1996. Photogrammetrie, Band2, Verfeinerte Methoden und Anwendungen. Duemmler Verlag, Bonn, 488 pages.

Scholten, F. and Gwinner, K., 2003. Band 12 "Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation", Bochum, pp. 419-426.

Vozikis, G., Jansa, J. and Fraser, C., 2003. Alternative Sensor Orientation Models for High-Resolution Satellite Imagery. *Band 12 "Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation",* Bochum , pp. 179- 186.

Vozikis, G., 2004: Automatic Generation and Updating of Digital City Models using High-Resolution Line-Scanning Systems. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34, Part XXX.

Vozikis, G. and Jansa, J., 2008. Advantages and Disadvantages of the Hough Transformation in the frame of Automated Building Extraction. International Society for Photogrammetry and Remote Sensing XXIst Congress, Beijing, China; 07-03-2008 - 07-11-2008; in: "Proceedings", Vol. XXXVII. Part B3b (2008), ISSN: 1682-1750; 719 - 724.

Zhang, Y. (2001): A Spectral and Spatial Information Integrated Approach for Tree Extraction from High-Resolution Digital Imagery. Digital Earth, Fredericton, Canada, 9 pages.