

AUTOMATIC BUILDING EXTRACTION FROM AIRBORNE LASER SCANNING DATA

Michel MORGAN*, Klaus TEMPFLI**

*Ohio State University, USA

Civil and Environmental Engineering and Geodetic Science Department

Morgan.465@osu.edu

**International Institute for Aerospace Survey and Earth Sciences (ITC), The Netherlands

Tempfli@itc.nl

TECHNICAL COMMISSION III

KEY WORDS: Laser scanning, building extraction, morphological filter, connected component labeling

ABSTRACT

Laser scanning is a new technology for obtaining Digital Surface Models (DSM) of the earth surface. It is fast method for sampling the earth surface with a high density and high point accuracy. In this paper a procedure for building detection and roof extraction from the DSM is presented. The procedure starts by re-sampling elevation as obtained by laser scanning into regular grid. The core part of the building detection is based on a morphological filter for distinguishing between terrain and non-terrain segments. The non-terrain segments are classified into building or vegetation. Aiming at a vector representation of buildings the roof faces are obtained by further segmentation of the building segments into sub-segments. The 3D geometrical properties of each face are obtained based on plane fitting using leastsquares adjustment. The reconstruction part of the procedure is based on the adjacency among the roof faces. Primitive extraction and face intersections are used for roof reconstruction. Prior knowledge about the buildings and terrain is needed for both the detection and extraction processes. The procedure is developed to work for all terrain types and for many building/roof types. The laser data used in this research have an average density of 2-3 points per square meter and 0.10m standard deviation of the elevation values. The procedure shows promising results for building detection. The reconstruction part shows promising results for some roof faces, obtaining high planimetric and height accuracy. Some adaptations of the procedure are recommended for enhancing the performance of the presented approach.

1 INTRODUCTION

Many applications, such as urban planning, telecommunication, security services ask for 3D city models. Buildings are the objects of highest interest in 3D city modeling. Urban areas are usually rapidly changing mainly due to human activities in construction, destruction or extension of topographic elements such as buildings and roads. This leads to requesting a fast data acquisition technique and automatic method for detecting and extracting 3D topographic objects from the data. Airborne laser scanning is a new technology in which several sensors are integrated to obtain 3D coordinates of points on the earth. It makes use of precise GPS instruments to determine the position of the sensor, inertial navigation system (INS) to determine the attitude of the sensor, and narrow laser beams to determine the range between the sensor and the target points. Laser scanning systems are active, therefore they can work day and night. Shadows also do not affect laser data.

In this paper an automatic procedure is presented for building detection and roof extraction from the DSM as obtained from laser scanning. Complete reconstruction of buildings with vertical walls from reconstructed roofs can be done following the approach described in (Tempfli, 1998). The laser data used in this research are described in section 2. The procedure for building detection and extraction is described in section 3. In section 4, empirical tests and the analysis of the results is presented. Section 5 includes the conclusions and the recommendations and the possible enhancements of the procedure.

2 DATA DESCRIPTION

The laser data used in this research are of an area located in Haren in The Netherlands obtained by TopoSys laser scanner. The data were kindly provided by the Surveying Department of the Ministry of Transportation, Public Works

and Water Management Directorate, The Netherlands. We obtained a standard deviation of the elevation values of 0.10 meters by analyzing the data. The average density (resolution) of the DSM is 2-3 points per square meter. The ground truth data that we used for reference is a vector map (lines) of the buildings, parcels, and roads in the area and check points with 3D coordinates. For the sake of exploring the data, we first produced shaded relief and contour maps from the DSM. The buildings can be easily seen in both maps. It has to be mentioned that we had no complete 3D data for the buildings at our disposal for a comprehensive evaluation of the results. Moreover, the accuracy and the degree of reliability of the ground truth data were not known.

3 THE BUILDING DETECTION AND EXTRACTION PROCEDURE

3.1 Data Re-sampling

Processing of irregularly distributed points as obtained from laser scanning takes more time than processing of regularly distributed ones. The suggested procedure starts by re-sampling the “laser data” into a regular raster, i.e., producing a two-dimensional array of elevation values. To this end the cell size and the interpolation method should be chosen sensibly.

3.1.1 Cell size

Unless many points are eliminated in the filtering process of the raw laser data the DSM from laser will have a fairly regular distribution of points. For a regularly distributed set of points, the cell size can easily be chosen such that each resulting cell contains only one point, thus avoiding information loss and redundancy. The cell size must be equal to the reciprocal of the square root of the “point density”. We have applied this formula to our data, considering the highest density available in our original DSM (3 points/m²). It has to be mentioned that the regular raster has the same values of the regular grid whose grid points are located in the centers of the raster cells. In this research, when analyzing elevation values and the elevation difference the grid points are considered, and when dealing with areas, we consider the regular raster, assuming constant elevation throughout a raster cell.

3.1.2 Interpolation method

Interpolation of the DSM from laser scanning points using linear or higher order degree of surface curvature gives a chance to have erroneous height values of the cells along the boundaries of the non-terrain objects, which leads to some difficulties in building detection. Therefore nearest neighbor interpolation is used, although it leads to shift in the building boundaries. However, as the grid size becomes smaller, this error will be smaller. Another drawback of using nearest neighbor interpolation occurs for inclined roof faces where the interpolation error has horizontal and vertical components. Therefore, two interpolation methods will be employed for the same data using the same grid size; nearest neighbor interpolation will be used for building detection and linear interpolation for the classification between buildings and vegetation and for building extraction.

3.2 Building Detection

3.2.1 Distinction between terrain and non-terrain segments

3.2.1.1 Morphological filter

Separation between terrain and non-terrain pixels is done by applying a morphological filter (Hug, 1997), (Hug and Wehr, 1997) and (Kilian et al, 1996). Using the morphological filter, different sizes of moving windows are used. For each window, the “deepest pixel” and all other pixels which are “higher” than the deepest pixel within a certain range (band width) are detected. Weights are assigned to these pixels according to the chance of being terrain pixels. The weights of the pixels in a large window should be larger than those in smaller windows. Moreover, among the deep pixels, the deeper the pixel, the larger its weight. Equation (1) is the formula for computing weights. The idea behind this equation is to scale the weights from 0 to 1 linearly between the smallest and the largest window sizes and also along the band width above the deepest pixel inside the window. It has to be mentioned that the morphological filter is used to threshold the DSM data in a window relatively smaller than that of the whole data set in order to compensate for the sloping/variable terrain.

$$\text{Pixel_weight} = \frac{\text{window_size} - \text{min_window_size}}{\text{max_window_size} - \text{min_window_size}} \cdot \frac{\text{band_width} - (\text{point_ht} - \text{deepest_point_ht})}{\text{band_width}} \quad (1)$$

Prior knowledge about buildings is used in determining the window sizes. The minimum and the maximum size of the windows used are selected to be slightly larger than the expected minimum and maximum building size respectively. The maximum and the minimum building size can be chosen according to the specifications of the application. The band width is the expected range of terrain elevation in the window. Some factors play a role in band width selection such as the terrain variability, the window size and the minimum building height. As the terrain roughness (or the standard deviation of the terrain elevation) increases, the band width increases. Moreover, as the window size increases the band width increases as far as the terrain has a certain slope. The band width has to be smaller than the minimum building height in order not to classify building roofs as terrain. All pixels should be classified by thresholding the weight values to distinguish between terrain and non-terrain pixels.

3.2.1.2 Connected component labeling

Segmentation is applied after having identifying the non-terrain pixels. Segmentation is based on the adjacency relationships among pixels. For this reason, connected component labeling is used to identify the connected pixels (based on 8 directions) and assign a unique label for each segment. To reduce of the noise and the misclassification, median filtering is used before connected component labeling. Also median filtering is used to eliminate thin lines and to disconnect some wrongly connected segments. The resulting segments are candidates to be building segments and will be further classified as described in the following section.

3.2.1.3 Distinction between building and vegetation segments

The distinction between building and vegetation segments is done based on the minimum building area and the change in elevation and slope/orientation within each segment.

Minimum building area

Using the minimum building area as a threshold value for the areas of the segments, we exclude and eliminate the small segments. Concerning a certain application, the threshold can be set according to the specifications of the required buildings. For example, if the building area is specified as larger than 50 m² to have a certain function as housing, this value is used as the threshold. Moreover, the maximum building area can be used also as another threshold.

Change in elevation

The change in elevation will be obtained by applying a Laplacian filter—with coefficients as presented in equation (2)—to the DSM. Within only the planer building faces, the pixel values in the resulting image are (ideally) zeros, while for vegetation and along edges the pixel values differ from zero. Segment shrinkage of one pixel width is done to exclude the boundary pixels because they disturb the classification.

$$\text{Laplacian Filter Coefficients} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (2)$$

The classification of the whole segment into building or vegetation is difficult due to the contribution of the edge pixels inside the building segment (i.e. the pixels between building faces). Therefore, the classification will be done pixel by pixel inside each segment. It is assumed that errors of the laser data are random errors and have a vertical component only. Moreover, it is assumed that there are no errors due to the interpolation (using bilinear interpolation). The standard deviation of the change in elevation is computed based on the coefficients of the Laplacian filter by applying the error propagation rule. A threshold value can be chosen based on the standard deviation of the elevation change. Each pixel will be classified in a binary classification into face pixel or vegetation (or edge) pixel. By counting the number of face pixels in comparison to the total number of pixels for each segment, the segment will be classified into building or vegetation. For example, if the number of interior face pixels is larger than 50% of the total number of pixels in the segment, the segment is classified as building. The selected ratio (threshold) depends on the face size, the size of chimneys, and the pixel size. Prior knowledge about building faces is required for selecting this threshold value. It has to be mentioned that the previous ratio can be used in determining whether the segment contains a building and adjacent vegetation detected as one segment in the building detection process.

Change in slope

The slope (or the inclination) and the orientation (or the azimuth) are computed from the DSM using the Sobel operator in two perpendicular directions. The standard deviation of the slope and orientation for each pixel is determined based on the filter coefficients used and applying the error propagation rule. By applying the Laplacian filter to the slope image, the values of the resulting image should be (ideally) zeros for the pixels inside the planar roof faces. As mentioned before, segment shrinkage is used to avoid the contribution of the boundary pixels in the classification, and the classification is done pixel by pixel within each segment. Again, the standard deviation of slope change is computed by applying the error propagation rule. The threshold value of each pixel in the resulting image can be defined based on the standard deviation of the slope change. Each pixel is then classified as face pixel or vegetation (or edge) pixel. It is again a binary classification. By counting the number of face pixels in comparison with the total number of pixels for each segment, the segment is classified into building or vegetation. This threshold can be calibrated based on the prior knowledge about the face area, chimneys size, and the pixel size.

It has to be mentioned that classification of flat buildings based on the orientation will fail because the orientations of the pixels of the flat buildings have random-like values. Therefore, a prior knowledge about the building roofs should be available to select the best criteria for classification. However, not considering the change in orientation in the segment classification is more conservative.

3.3 Building Extraction

3.3.1 Identifying roof faces

3.3.1.1 Connected component labeling

The strategy for identifying building faces is to connect adjacent pixels (based on 8 directions) of the same slope and/or the same orientation. Because the slope and orientation values contain errors, threshold values should be considered for the change in slope and orientation between the adjacent pixels. The threshold value can be defined based on the standard deviation of the difference of the slope and orientation between the adjacent pixels. Pixels are then segmented into sub-segments within the building segment.

Having more than one criterion for connecting pixels leads to better results (as comparable with having many bands in image classification). However, connecting adjacent pixels based on the orientation will lead to unexpected results in the case of flat roofs. Therefore, connecting adjacent pixels is used based on slope and orientation for sloping roofs, and based on slope only for flat roofs. Therefore, prior knowledge about the building roofs is used to select the criteria for the segmentation. However, if the pixel slope value is larger than a certain value (theoretically zero) both slope and orientation have to be used, otherwise only the slope has to be used. The mentioned value is chosen based on the standard deviation of the slope or based on prior knowledge about the minimum slope of the sloping roofs. For example, if the minimum slope of the sloping roofs is 30° , a value of 15° can be used as long as the standard deviation of the slope is less than 15° . The suggested method for connecting adjacent pixels is meant to work also in the cases where horizontal and sloping faces are contained in the same building.

3.3.1.2 Majority filtering

After applying the connected component labeling, several sub-segments within the building segment are identified. To correct for the small sub-segments, which may be caused by the chimneys or other “disturbances”, a majority filter is used. The size of the window of the majority filter can be chosen based on the knowledge about the chimney size and the pixel size. Alternatively an iterative majority filter of small window size (e.g. 3 pixels by 3 pixels) can lead to the same result without this pre-knowledge. Iterated majority filtering is then applied till stability is reached, i.e. till no more changes happen. It has to be noted that the majority filter has to be constrained to make sure that the building boundaries do not shift. In other words, we have to keep the building segment unshrunked and unexpanded. Face segments before applying the majority filter are assumed to have only random errors in the height values. After applying the majority filter more non-random errors (in the height values) in the reclassified pixels will exist. The height values of those reclassified pixels will not be considered for extraction of the height and slope information of the faces. However they are used for extracting planimetric information. Another advantage of the mentioned reclassification is to determine the adjacency information among roof faces.

3.3.2 Face extraction/ vectorization

A least squares adjustment for plane fitting is used based on the height values of each of the main roof faces before applying the majority filter (to make sure that there are no gross errors or blunders in the height values). After applying the iterative majority filter, the adjacency among the roof faces can be obtained. The way of vectorizing the face depends on the adjacency among faces.

If the face is adjacent to only one or no other faces in addition to the background, primitive matching (matching the face segment with one of the geometric primitives and extract the primitive parameters) is done for that face. The geometric primitives can be chosen according to an inventory of roofs in the area of interest. In this case the planimetric information is used to obtain the primitive parameters and the height information is added to the primitive based on the plane equation of the face. Primitive matching can be done based on the analysis of the compactness and the relative moments of area along the principal axes of the face. However, because of the time limitations, primitive matching was not used in the present research. Instead, only rectangle extraction is used based on defining the principal axes and matching the moments of area along the principal axes with the parametric rectangle, followed by extracting the rectangle parameters. Model matching had been done before in (Maas, 1999), (Weidner and Forstner, 1995) and (Weidner, 1996). However, in this research matching is done for each roof face to consider more complicated roof structures.

If the face is adjacent to more than one face, the outline of the face are extracted based on the intersection of the face and the adjacent face. For the faces adjacent to the background line fitting is done to obtain the respective edges.

4 EXPERIMENTS AND RESULTS

The laser data is re-sampled into regular raster with a pixel size of 0.577 m which is the invert of the square root of the maximum density of the laser data (3 points per square meter). The data are re-sampled twice; using nearest neighbor interpolation, which is used for building detection and using bilinear interpolation, which is used for differentiating between buildings and vegetation and for building extraction. The procedure for building detection and extraction is done for two test areas. The DSMs of the test areas after re-sampling into regular raster are shown in figure (1). The DSM is represented as an image, elevation values converted to gray values.

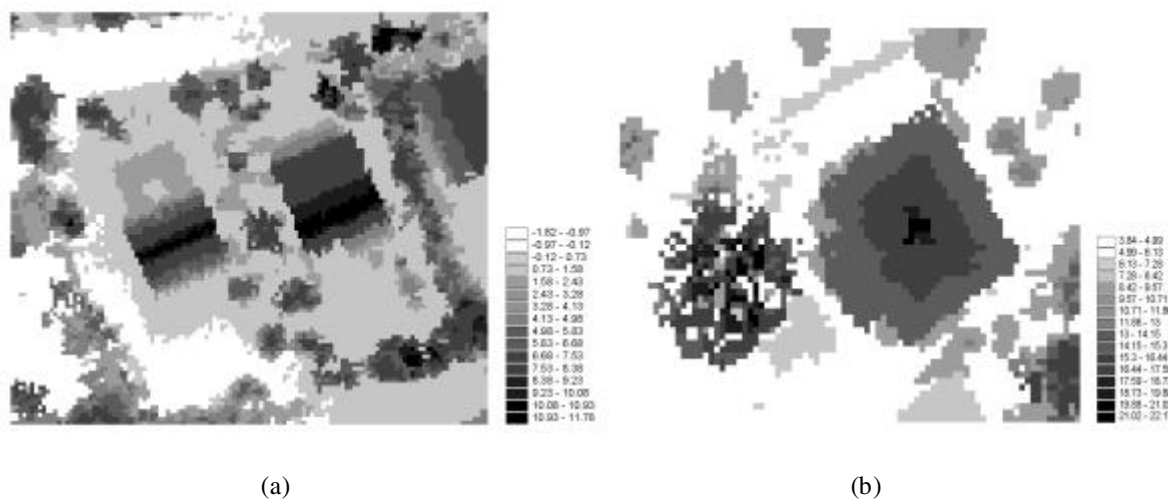


Figure 1. The DSM after re-sampling the laser data into regular raster for (a) the first and (b) the second test area

Applying the median filter before the connected component labeling re-classifies some of the misclassified pixels or noise inside/outside the objects. Moreover some of the thin lines connecting segments are eliminated. The elimination of those thin lines is important in the case where 8-directions connectivity is used for segmentation. The number of segments is smaller than that without applying the median filter. The minimum building area used for classification is

50 m². A threshold of 50% is used for the ratio between the interior face pixels and the total number of pixels for each segment.). Two building segments are detected for the first test area and one for the second test area as shown in figure (2). Buildings are successfully detected according to the visual interpretation of the DSM and the global comparison to the ground truth map.

Three and two main faces are identified for the buildings in the first test area and four main faces for the building in the second test area. Because one of the faces of the first building is outside the ground truth map we excluded it from the building extraction processes. The next step is to obtain adjacency information among the roof faces. The outcome for the first test area is that each face is adjacent to one face in addition to the background. Each face is adjacent to two or three other faces in the second test area. Therefore, rectangle extraction is done for each of the two faces of the buildings in the first test area. A height value is assigned to each corner of each rectangle according to the equation of the face plane and the planimetric coordinates of the corner. For the building in the second test area, face intersections based on the plane equations are done for adjacent faces and line fitting is done for the boundary pixels for the faces adjacent to the background. The outline vertices of the building could be calculated by intersecting the outlines. Considering the line intersections between faces are more reliable, the outline vertices are obtained by intersecting the lines (which are the intersections between the faces) and the outlines. This leads to two solutions for each vertex and by averaging their coordinates one solution is obtained for each vertex. The resulting roof faces are overlaid with the ground truth map and the detected building segment and displayed in figure (3).

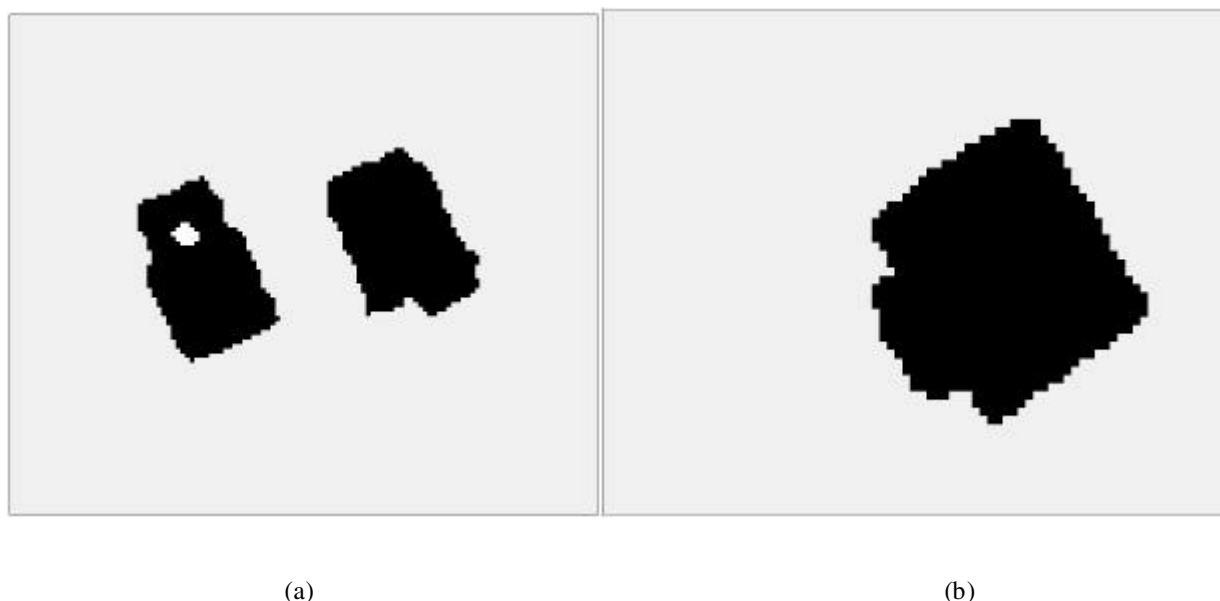


Figure 2. The detected building segments in (a) the first and (b) the second test area.

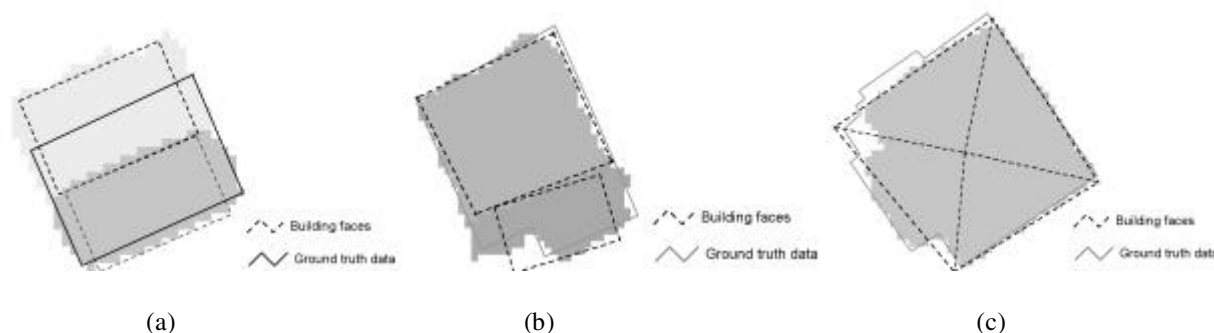


Figure 3. The resulting roof faces overlaid with the detected building segments and the ground truth data for the buildings in (a and b) the first and (c) the second test area.

For the building in figure (3-a), the direction of the extracted building is the same as the direction of the building in the ground truth data but the dimensions of each rectangle differ from the ground truth values. The two extracted faces of the building slightly overlap within the pixel size. This building is on the edge of the ground truth data and it was not known whether the building in the ground truth data is trimmed at the upper part of the figure. Moreover, the ground truth data does not have face information. Therefore, only the lower left and lower right coordinates of the building in the figure are used for accuracy estimation. The planimetric and the height accuracy are 2.25m and 0.16m respectively. The accuracy achieved for the height (0.16m) is very good compared to the standard deviation of elevation from laser scanning (0.10m). The planimetric accuracy (2.25m) is very low in comparison to the pixel size (0.577m). The results of building detection (the gray pixels in figure (3-a)) are better than the results after vectorisation (the dotted lines in the same figure). As shown in the same figure the extracted faces are smaller than the face segments in raster format. Area constraint for each face may be used as a constraint in the rectangle extraction process (hopefully to extract the correct rectangle dimensions). Moreover, the rectangle extraction could be constrained to the direction of the intersection line between the two planes of the faces.

For the second building, the extracted rectangle of the upper face in the figure (3-b) is rotated by 4 degrees with respect to the ground truth data. Again because there is no face information in the ground truth data, only the upper right and upper left corners of that face will be checked. The planimetric and the height accuracy are 0.55m and 2.47m respectively. The achieved planimetric accuracy (0.55m) is very good and it is within the pixel size (0.577m). The very large error in height is caused by the erroneous plane parameters of the faces. The segmentation of the building segment into sub-segments (faces) is too sensitive to noise because the criteria for segmentation are applied for pairs of adjacent pixels. Therefore, the criteria for segmentation must not be applied only for each two adjacent pixels while doing region growing, but also for the pixel and the growing sub-segment. The second face in the building is not rectangular as can be seen in the figure, therefore the rectangle extraction results in a rotated rectangle with different dimensions than that in the ground truth data. Constraining rectangle orientation to the direction of the intersection between the faces could lead to better results for the second face.

For the building shown in figure (3-c), the boundaries of the extracted faces are generalized because only one line fitting is done for each face adjacent to the background. For the four main building corners, the planimetric and the height accuracy are 0.73m and 1.30m respectively. The planimetric accuracy (0.73m) is 1.3 times the pixel size (0.577m). The reason behind achieving a planimetric error larger than the pixel size is the generalization of the building outlines. The low accuracy in height is obtained because of the erroneous plane parameters of the face planes due to the face segmentation process. Moreover, the low planimetric accuracy results in low height accuracy having sloping faces. Model refining/recovery can be done to reconstruct the building (hopefully) in order to improve the results.

5 CONCLUSIONS AND RECOMMENDATIONS

In this research a procedure for building detection and roof extraction from a DSM obtained by laser scanning is investigated. The procedure is meant to work for all terrain and many building types. Empirical testing, however, was only done in two areas of low terrain variability. Multi level roofs are not considered in the procedure, which could be modified to consider them. The procedure shows promising results for the building detection part and for some roof faces in the extraction part. We have only considered planar faces, therefore, curved faces can not be detected as one face. Prior knowledge about buildings and terrain type is required for the developed building detection and extraction.

More experiments are recommended, including test areas of high terrain variability. More calibration studies on the threshold values should be done. In the suggested procedure, building reconstruction is dependent on the results of the building detection. Moreover, building pixels are not allowed to be reclassified into non-building pixels and vice versa. Therefore, vegetation which is adjacent to a building will be detected and constructed as a part of the building. Analysis of the subsegments inside a large segment (or object), in the same way as the classification of objects into buildings or vegetation, has to be done in order to single out the vegetation and reconstruct the building using only the roof faces. Some adaptations of the segmentation of the face segments should be done in order to reduce the effect the noise in the segmentation process. More work in primitive matching and extraction has to be done as well. More constraints can be added to the primitive extraction such as the face area and the intersection line between the adjacent faces. In the used procedure the laser data are resampled into a regular grid, which can lead to geometric errors. It has to be investigated if the procedure could be modified to directly use the original irregularly distributed data. Once the procedure is refined it would also be of interest to study its performance on DMSs originating from other data sources.

ACKNOWLEDGEMENTS

We would like to acknowledge the support of the Surveying Department of the Ministry of Transportation, Public Works and Water Management Directorate, The Netherlands by providing the data.

REFERENCES

Hug, C., 1997. Extracting Artificial Surface Objects from Airborne Laser Scanner Data. Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) Monte Verita, 1997.

Hug, C. and Wehr, A., 1997. Detecting and Identifying Topographic Objects in Imaging Laser Altimeter Data. IAPRS, Vol. 32, Part 3-4W2, "3D Reconstruction and Modeling of Topographic Objects", Stuttgart, September, 1997.

Kilian, J., Haala, N. and English, M., 1996. Capture and Evaluation of Airborne Laser Scanner Data. International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, Vienna, 1996.

Maas, H., 1999. Fast Determination of Parametric House Models from Dense Airborne Laserscanner Data, ISPRS Workshop on Mobile Mapping Technology, Bangkok, Thailand, April 1999.

Tempfli, K., 1998. 3D topographic mapping for urban GIS. ITC Journal 1998-3/4, pp.181-190.

Weidner, U., 1996. An Approach to Building Extraction from Digital Surface Models. 18th ISPRS Congress, Wien, Proceedings B3, pp. 924-929.

Weidner, U. and Forstner, W., 1995. Towards Automatic Building Extraction from High Resolution Digital Elevation Models. ISPRS Journal, Vol. 50, No. 4, pp. 38-49.