RELIABLE RECONSTRUCTION OF BUILDINGS FOR DIGITAL MAP REVISION

Markus NIEDERÖST
ETH Zürich, Switzerland
Institute of Geodesy and Photogrammetry
markus@geod.baug.ethz.ch

Working Group III/4

KEYWORDS: Buildings, Blob detection, Classification, Vector data, Reconstruction, Map revision

ABSTRACT

This paper describes a framework for automatic reconstruction of buildings in order to correct and update an initial 2-D vector data set and to derive a 3-D model for visualization. The color images of a stereo model are used to calculate a digital surface model and an orthophoto. A procedure for blob detection from the calculated digital surface model as well as unsupervised multichannel classification are used to produce additional approximate vector data. This allows to introduce buildings to the data set that were not yet included. Methods of digital image processing are used to improve planar location, rotation and scales of the vector data for each single building. The solutions are rated in order to keep the good results and to reject inaccurate or wrong data automatically. Finally one representative height for each building is derived from the digital surface model. The accuracy of the building reconstruction is determined by quantitative comparison with reference data, and the rating procedure is checked. The described system neither requires large image scales nor threefold up to sixfold overlap nor laser scanner data nor infrared information to automatically revise the initial vector data and to produce a coarse 3-D city model. Approximation vector data can be used, but it is not mandatory.

1 INTRODUCTION

The work presented in this paper is part of the project ATOMI (Automated reconstruction of Topographic Objects from aerial images using vectorized Map Information), a cooperation between the Swiss Federal Institute of Topography, Bern (L+T) and the Institute of Geodesy and Photogrammetry (IGP) at ETH Zürich. Primary aim of the project is to contribute to the production of a 3-D landscape model of Switzerland including buildings and roads. Input data to be used are aerial images which are produced for the whole area of Switzerland in a 6 years cycle. Approximation vector data is acquired by L+T from digital raster maps (scale 1:25'000) by means of semi automatic vectorization procedures. The questions that are being investigated at IGP are - for roads and buildings - to automate the refinement and revision of the vector data. Besides the improvement of accuracy this also includes the determination of one representative building height and the completion respectively update of the vector data set. For a general overview of the project see (Eidenbenz et al, 2000), the part about roads is described in (Zhang/Baltsavias, 2000).

Reconstruction of buildings is currently one of the main topics of the photogrammetric community, and the approaches to solve the problem as well as the used kind of data are manifold. (Henricsson, 1996) used fourfold overlapping colour images (pixelsize in object space 7.5 cm). A procedure with even sixfold overlap is described in (Baillard et al, 1999) where the ground resolution is 8.5 cm. Both approaches take profit of the high ground resolution and the number of images and thus reduce the problem of edges that are not contained in some of the images. (Weidner, 1997) or (Maas, 1999) use very precise digital surface models from airborne laser scanning systems, while the classification procedures of (Haala/Walter, 1999) base on color infrared images which simplify the separation of man-made objects from vegetation. An overview of various approaches can be found in (Grün, 1995) and (Grün, 1997).

In the project ATOMI neither large scale images with high resolution or infrared images are used nor is the digital surface model produced by laser scanner or radar. Due to economical reasons the procedures being developped have to reliably work in regions with twofold overlap. On the other hand a priori knowledge (approximate vector data) can be used. Aim of the work described in this paper is not to get a fully automatic system but a very reliable system. Buildings that cannot be automatically reconstructed with sufficient precision have to be automatically rejected and can be remeasured manually.

2 DATA

2.1 Image data

The used test region is the village Hedingen south of Zürich. The color images of one stereo model (mean flight height above ground 4'800 m, focal length ~ 300 mm) were scanned with a resolution of 28 microns. At an image scale of 1:15'800 this resulted in a ground resolution of 0.45 m.

2.2 Digital terrain model

The digital terrain model (DTM) was provided by the Swiss Federal Institute of Topography (L+T). This so-called DHM25 is derived from the contour lines of topographic maps (scale 1:25'000) through interpolation. It is available for the whole area of Switzerland with a rasterwidth of 25 m. The accuracy of the DHM25 is around 1.5 m for the Swiss Plateau and approximatively 5 to 8 m for the alpine area.

2.3 Digital surface model

Commercial software (Phodis by Zeiss) was used to orient the stereomodel and to generate a digital surface model (DSM). Several tests have been made to find an optimum parameter set which preserves the surface shape including buildings as accurate and detailed as possible (Fig. 1). A comparison of DSMs derived from 14 µm images and from 28 µm images and the choice of a rasterwidth of 2 m or 1 m respectively showed no significant difference of the results. The better resolution and smaller gridwidth even resulted in disturbing small details and height errors. For reasons of saving processing time and disk space the 28 µm images have been used and a DSM rasterwidth of 2 m was chosen. In addition not the whole overlapping region was used but the test area to be processed was reduced to an extent of 500 m x 440 m.

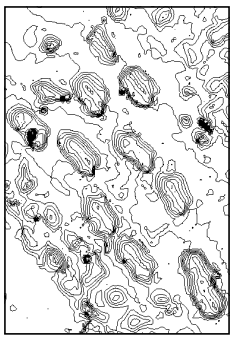

Fig. 1: Part of DSM

2.4 Orthophoto

A color orthophoto of the test area (Fig. 3) with ground resolution 0.25 m was calculated using one image of the stereomodel and the previously produced DSM. The image data for multichannel classification and house reconstruction was derived from this orthophoto.

2.5 Approximate 2-D vector data

The vector data set (VECTOR25) was produced at L+T by semi-automatic vectorization from digital maps 1:25'000. Characteristics of the initial vector data relevant for this project are:

- Level of detail is according to the content of the 1:25'000 map (generalized)
- 2-D (no height values)
- New buildings are not yet included
- Old buildings which don't exist anymore are still included

3 DETERMINATION OF APPROXIMATE BUILDING LOCATIONS

The initial location of buildings is determined both by approximate vector data as well as by two procedures for building detection. The detection is essential because houses that are not yet included in the vector data set have to be added. All three resulting data sets - VECTOR25, result from blob detection, classification result - are used as initial data for the building reconstruction.

3.1 Vector data

Each building in the approximation vector data set provides one building approximation.

3.2 Blob detection

The blob detection from height data is done in 3 steps carried out either using the DSM or the normalized DSM (difference between DSM and DHM25) as input data. The existence of the DHM25 (see '2.2 Digital terrain model') would also

allow simpler techniques like a thresholding in the normalized DSM. On the other hand the procedure described below is also applicable in projects where only the DSM is available and thus more general.

3.2.1 Calculation of height bins: The height information in the format of a regular grid of z-coordinates is cut into binary height bins, each of them corresponding to one height interval. The chosen height difference between two bins is 1 m. Generation of height bins with a thickness of 1.5 m results in an overlap between adjacent bins to avoid problems in the subsequent steps caused by inaccuracy of the DSM or the normalized DSM respectively. Each rasterpoint in a bin equals 1 if its z-coordinate lies in the height interval of the current bin, otherwise the value is 0 (Fig. 2).

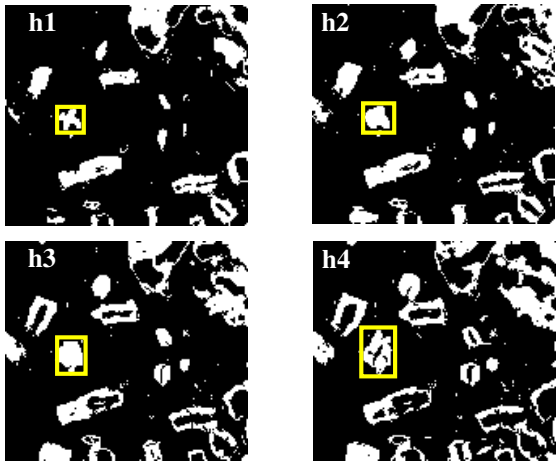

Fig. 2: Four consecutive slices for a part of the test area (heights h1>h2>h3>h4) and enclosing boxes for one building

3.2.2 Find elements in bins: The bright objects consisting of a group of adjacent white pixels bounded by black ones are detected and for each element an enclosing rectangle (Fig. 2) as well as the centre of this rectangle are determined, providing one list of elements ordered from the highest to the lowest bin.

3.2.3 Tracking along z: In order to detect the blobs, the list of elements is processed from the highest bin down to the lowest bin. If through a sequence of several height bins a chain of elements can be found at the same location (distance of centres of gravity < 7 m in object space), a possible blob is detected and described by the second highest enclosing rectangle of the chain. The size of this rectangle approximatively corresponds to the building dimensions.

Main problem of the blob detection is that buildings cannot be separated from vegetation. Furthermore buildings standing close to each other or trees standing too close to buildings are detected as one blob.

The blob detection results in one list of northwards oriented rectangles as input vector data for the house reconstruction.

3.3 Multichannel classification

The two color channels R (red) and G (green) from the RGB color space, the channel a* (redness-greenness) and L* (brightness) of the CIELAB color space and the channel S (saturation) of the HSI color space have been used to calculate several channels for the classification procedure. The channels of CIELAB and HSI can be derived directly from the RGB color space (Pratt, 1991). Adding one channel containing the normalized DSM, the following channels were used:

- Channel containing shadows, derived from the S channel of the HSI color space by thresholding
- Channel a* from the CIELAB color space
- Channel containing texture. Texture means the number of edge pixels in a circle around the centre pixel (diameter 8 m) as proposed in Braun (1999).
- Channel containing the degree of artificiality (DoA). The result of (G-R)/(G+R) was first inverted and then a histogram equalization was done. The brighter pixels belong to man-made objects (high grade of artificiality), while darker pixels are part of vegetation (see Fig. 5).
- Channel containing the normalized DSM (nDSM)

All channels were stretched to the interval [0, 255]. The classification which is an improvement of the procedure described in (Braun, 1999) consists of the following steps:

- Removal of shadows (using channel S). Regions with shadows are excluded from further processes.
- Unsupervised K-means classification (2 classes) to separate above-ground regions from ground regions (input channels a* and nDSM). Regions classified as ground regions are excluded from further processes.
- Unsupervised K-means classification (2 classes) of above ground regions to separate man-made objects from vegetation regions (input channels texture and degree of artificiality). Regions classified as vegetation regions are excluded from further processes.
- Postprocessing to remove small objects and to fill small gaps

Fig. 3: Orthophoto of the test area

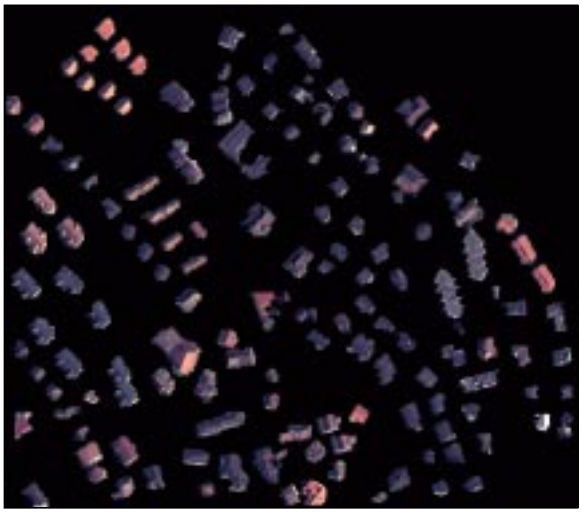

Fig. 4: Classification result

Application of the multichannel classification procedure to the test area (Fig. 3) produced a result which shows a clear separation of buildings from other image content (Fig. 4). Although trees close to buildings could not be completely eliminated and shapes of buildings are often not accurate, the result is sufficient for being used in further steps.

In order to provide data which can be used as approximate data for the reconstruction, the classification result is processed to get a list of bounding boxes, including one bounding box for each classified building. Those northward oriented rectancles serve as input vector data for the building reconstruction.

4 BUILDING RECONSTRUCTION

The building reconstruction process is done for each initial building seperately. Input for the reconstruction can either be the approximate vector data provided by L+T (VECTOR25), the result from the blob detection or a bounding box derived from the classification result.

The building reconstruction steps described below are done using image data based on the calculated color orthophoto and height information. Those are the degree of artificiality (DoA) described in '3.3 Multichannel classification', the absolute height values from the DSM, the normalized DSM and the L* channel (brightness) which is used to calculate orientation and magnitude of the grey value gradients using the Sobel operator (Fig. 5).

This new approach profits of the fact that a multiplication of the normalized DSM with the degree of artificiality results in a separation of artificial objects above ground from all other objects.

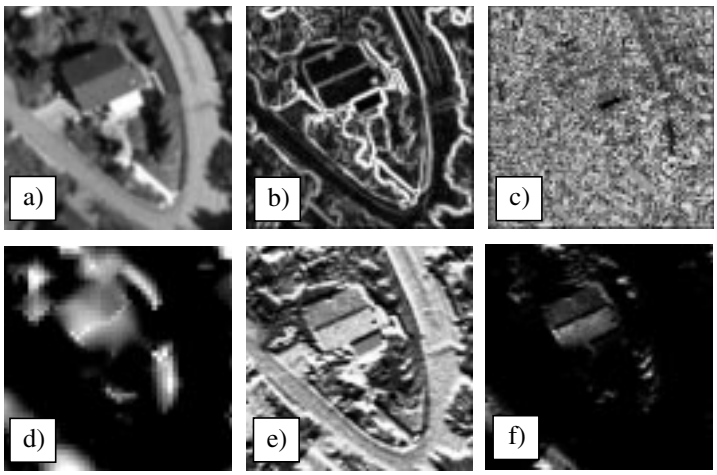

Fig. 5: Used data for building reconstruction:
a) L* channel (brightness)
b) magnitude of grey value gradients (from L*)
c) orientation of grey value gradients (from L*)
d) normalized DSM
e) degree of artificiality (from R and G)
f) multiplication of the normalized DSM with the degree of artificiality

4.1 Translation

First step of the building reconstruction is the improvement of the planar position. A binary mask (Streilein, 1999) calculated from the approximate vector data is moved on a regular raster (rasterwidth 2.5 m), going from -10 m to 10 m in x- and y-direction. On each of those 81 positions the following score (Eq. 1) is calculated:

$$Score = 2 \cdot \sum_{i=1}^{m} \frac{w_1 \cdot DoA \cdot nDSM}{m} + \frac{1}{3} \cdot \sum_{j=1}^{n} \frac{mag}{n} - dx - dy \quad (1)$$

m = number of pixels inside binary mask (= area), n = number of pixels under binary mask outlines (= lines),
w_1 = 2 - (distance of pixel to centre of mask)^2 / ((average extension of mask) / 2)^2,
dx, dy = absolute x- or y-coordinate difference (in pixels) between the actual mask location and the initial location given in the vector data,
DoA = degree of artificiality, nDSM = normalized DSM, mag = magnitude of grey value gradients

The weight w_1 is used to give a pixel in the centre of the binary mask a weight which is approx. = 1. Towards the mask outlines the weight decreases to approximatively zero. The values dx and dy are used to reduce the score depending on the distance between the new position and the initial position, assuming that the vector data location was almost correct. This is done to avoid that the vector data is translated to a building in the neighbourhood due to higher score. To allow faster processing times, the distance between two checked positions (2.5 m) is rather high. The caused inaccuracy will be corrected in '4.3 Scaling'. Finally the vector data is moved to the position with the maximum score (Fig. 6).

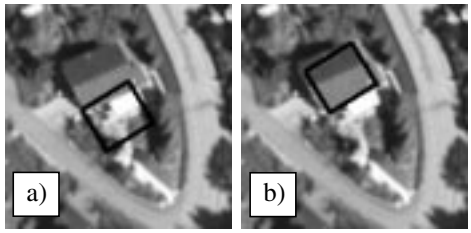

Fig. 6: a) approximate position
b) solution with maximum score

4.2 Rotation

Usually the direction of an edge in the orthophoto slightly differs from the direction of the adequate vector in the approximation data. This is relevant especially for the case when the approximation data is the result from either the blob detection or the multichannel classification, which always result in a square box oriented northwards.

For the calculation of the angle between approximation data and the building in the orthophoto the direction of the vectors is calculated and reduced to the interval [0, 89]. Using the Sobel operator on the L* channel the gradient direction and the magnitude of edges are calculated inside a mask which is enlarged with factor 0.4 to be sure to enclose all essential edges in case that the building in the vector data was too small (Fig. 7). Now the magnitude sum for each discrete gradient direction reduced to [0, 89] is calculated. The mean value of the two angles with the highest magnitude sum gives the correct orientation of the building (Fig. 8).

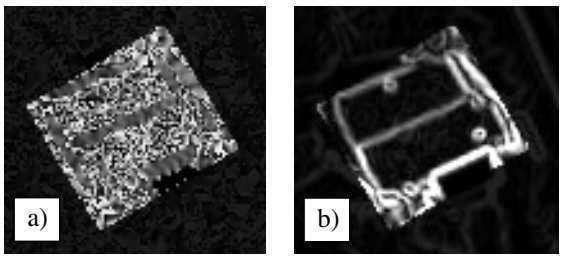

Fig. 7: a) gradient orientation inside of binary mask
b) magnitude values inside of binary mask

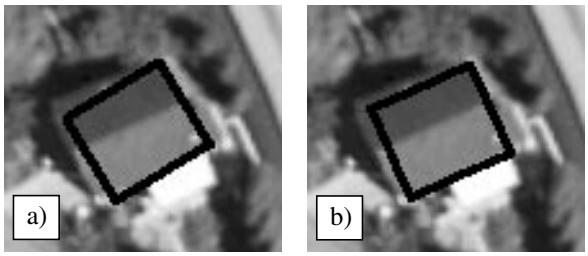

Fig. 8: a) before correction of orientation
b) result after correction of orientation

4.3 Scaling

The result after the correction of the orientation still does not exactly correspond to the building in the orthophoto. This is caused by the rough determination of the planar position. Additionally the size of the building in the vector data (and sometimes even the shape) is not the same as it is in the image data. For this reason an adaption of the scale(s) is necessary. The scaling (mask see Fig. 9) is done for the four orthogonal main directions of the vectors separately. On each side several positions closer to the building centre and further away are checked, and for each position a score is calculated (Eq. 2), which - similar as in '4.1 Translation' - makes use of the normalized DSM, the degree of artificiality and the magnitude of the edge gradients.

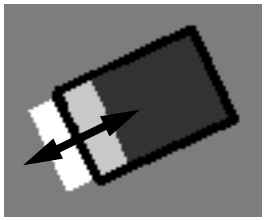

Fig. 9:
Mask for scaling (with black building outline)

$$Score = \sum_{i=1}^{m} \frac{DoA \cdot nDSM}{m} + 0.5 \cdot \sum_{j=1}^{n} \frac{mag}{n} + \left(\sum_{k=1}^{o} \frac{DoA \cdot nDSM}{o} - 1.5 \cdot \sum_{l=1}^{p} \frac{DoA \cdot nDSM}{p} \right) \quad (2)$$

m = number of pixels under the dark grey and the light grey area (inside building),
n = number of pixels under the black outlines (black),
o = number of pixels under the light grey area, p = number of pixels under the white area,
DoA = degree of artificiality, nDSM = normalized DSM, mag = magnitude of grey value gradients

The idea of the function is to have high values for artificiality and nDSM values inside the building and - described by the term in brackets - to have a high difference of (DoA · nDSM) between the adjacent regions inside and outside of the currently processed side of the building, which is usually the case for building edges.

The correction of scale is done for each side separately and just afterwards the new solution is calculated, using the corrections for the four sides (Fig. 10).

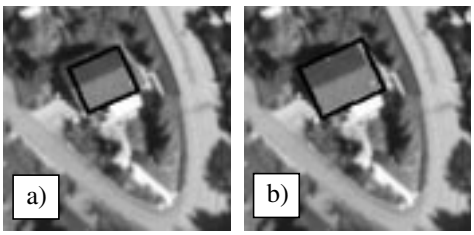

Fig. 10: a) before scaling
b) after scaling in 4 directions

4.4 Height

In order to get a height for each building the z-values from the DSM are used. The representative building height is calculated as the median value of all points whose planar coordinates lie inside the final building solution.

4.5 Rating of solutions

Aware of the fact that the procedures described above cannot produce the correct result for all processed buildings (no change of the initial shape characteristics, photogrammetrically derived DSM with possible errors, blobs caused by trees, buildings in the vector data that do not exist anymore) it is necessary to estimate the quality of the results using a score function. The products (DoA · nDSM) in term 1 and term 2 are clipped at the empirically determined cut-off value 25, in term 3 at value 100, allowing only lower values. This is done to damp the influence of very high values or outliers. The quotients are divided by the cut-off values to normalize each of the terms to the interval [0, 1]. With the average values under the 3 different areas (Fig. 11) the score can be calculated:

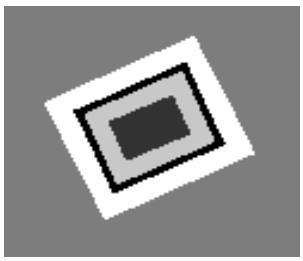

Fig. 11: Mask for rating of results (with black building outlines)

$$Score = \overbrace{0.5 \cdot \left(\sum_{i=1}^{m} \frac{DoA \cdot nDSM}{m} + \sum_{j=1}^{n} \frac{DoA \cdot nDSM}{n} \right)}^{Term\ 1} - \underbrace{abs\left(\sum_{i=1}^{m} \frac{DoA \cdot nDSM}{m} - \sum_{j=1}^{n} \frac{DoA \cdot nDSM}{n} \right)}_{Term\ 2} - \underbrace{\sum_{k=1}^{o} \frac{DoA \cdot nDSM}{o}}_{Term\ 3} \quad (3)$$

m = number of pixels under the dark grey area (inside building),
n = number of pixels under the light grey area (inside building),
o = number of pixels under the white area (outside building),
DoA = degree of artificiality, nDSM = normalized DSM, mag = magnitude of grey value gradients

Features of Equation 3 are that the height values and the artificiality outside of the building (white area) have to be small while the values inside the building (dark grey and bright grey area) have to be high. The difference between the averages of the two inside areas (term 2) has to be small, because a big value would mean that the final building solution includes high and low areas or man-made and natural objects.

Solutions with a score higher than 0.70 are automatically accepted. If the score is lower than 0.50, the solution is rejected and the building has to be measured manually. Buildings with a score between 0.50 and 0.70 have to be checked by an operator (Fig. 12).

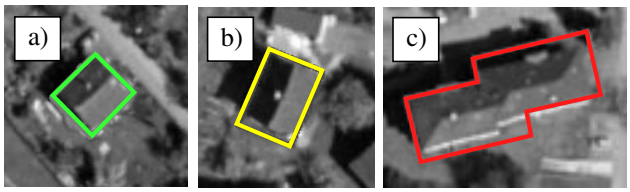

Fig. 12: a) accepted (score 0.79)
b) to be checked manually (score 0.69)
c) rejected (score 0.41)

Aware of the fact that with the used data (0.45 m ground resolution of original images, photogrammetrically derived DSM) and the applied techniques an accurate reconstruction of 100% of the buildings is illusory, it is essential to check if the building solutions are correctly rated to the three classes 'accepted', 'to be checked' and 'rejected'. The rating procedure automatically rejected bad solutions. The middle class contained the border cases which were classified manually. Most critical was the class 'accepted'. As no check by an operator is intended it is essential to avoid any errors of second order. No wrong solutions were erroneously classified to be correct.

5 EVALUATION OF THE COMPLETE FRAMEWORK

The described procedures were applied to the test area using the three kinds of information which are the vector data by L+T, the result from the blob detection and the enclosing boxes from the multichannel classification. Building reconstruction with all three data sets seperately led - in most cases - to several strongly overlapping solutions at each particular building location. Accepted solutions were automatically kept in the data set, bad ones were rejected. Uncertain buildings were visually checked on a digital photogrammetric system (Socet Set by LH Systems). Good ones were kept in the data set, bad ones were rejected. In a last processing step the best solution at each building position was selected, leading to the final data set. Missing buildings would have to be added manually. 3-D data for visualization is produced automatically (Fig. 13).

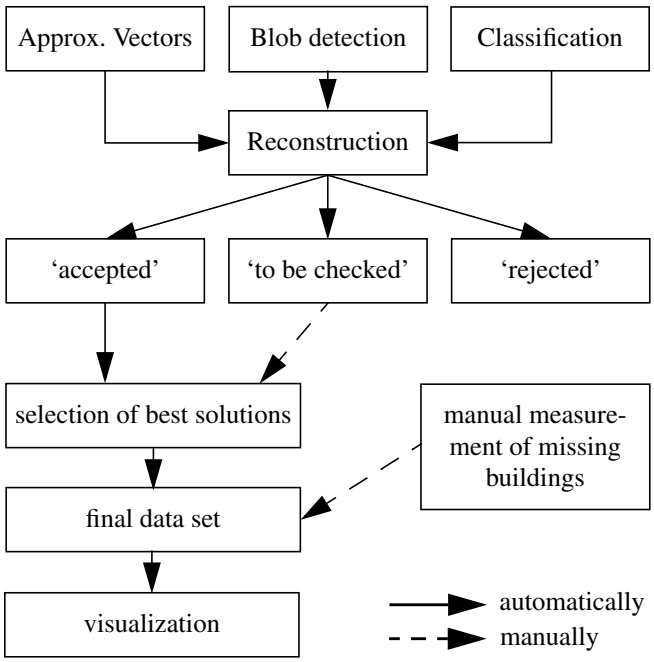

Fig. 13: Complete framework for detection and building reconstruction including addition of new buildings and removal of buildings that don't exist anymore with final conversion for 3-D visualization

The use of the 3 different kinds of initial vector data serves to reduce the probability that a building is not processed at all. Blob detection and multichannel classification allow to add new buildings that were not included in the approximation vector data. Using the scoring system it is also possible to remove houses that do not exist anymore. The vector data of such a building is processed, but it will finally be rejected due to a very low score.

buildings in test area	141 (100%)
buildings in vector data	117 (82%)
reconstructed buildings	126 (89%)

Tab. 1: Results for test area

The result achieved in the test area is promising. Counting all buildings in the test area the data processing ended up with 126 reconstructed buildings (89%). 19 houses were included that haven't been in the approximation vector data set. A mean accuracy of about 1 m for a single building side was quantitatively determined by comparison with manually measured reference data.

6 CONCLUSIONS

Although the described system produces a nice result, there is still potential for improvements. The processing time for building reconstruction could be significantly reduced if the vector data from blob detection or from multichannel classification would just be processed in the case that no correct solution (class 'accepted') was already produced from VECTOR25 at the same location. On the other hand multiple equal solutions at one location increase the reliability of the

result. A further improvement would be to use both aerial images, eliminating the influence of errors in the DSM during orthophoto generation and taking benefit of the twofold overlap.

A high potential lies in the use of different input data. Especially the use of a very precise DSM from laser scanner data and color infrared or infrared imagery of higher resolution would significantly improve the accuracy and reliability of each single processing step.

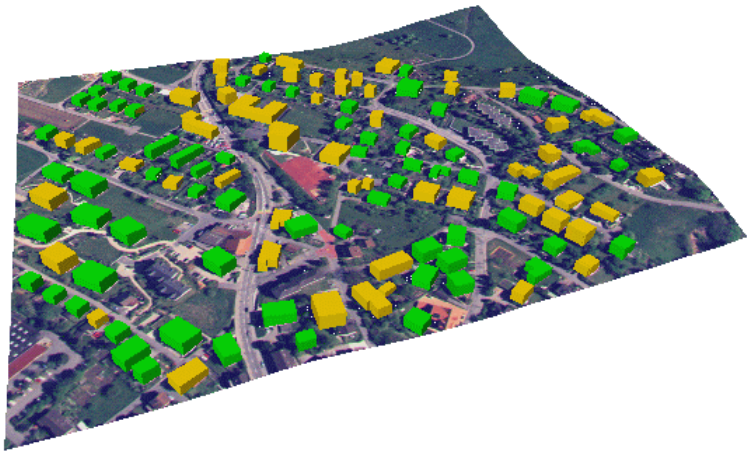

Fig. 14: Visualization of results with green (automatically accepted) and yellow buildings (manually accepted) without manual measurement of missing buildings

It was shown that the use of a single stereo image pair allowed to automatically revise the initial vector data set and to produce a coarse city model. An example (without manual measurement of missing buildings) can be seen in Fig. 14. Two procedures for building detection (blob detection from DSM, multichannel classification) were used to produce approximation data sets for those buildings that were not included in the initial vector data. The approach for building reconstruction allowed the correction of planar position, rotation and size of most of the buildings including the determination of a representative height. The produced solutions were automatically rated and data for 3-D visualization was generated. The main advantage of this whole framework is that input data can be restricted to one single image stereo pair.

REFERENCES

Baillard C., Schmid C., Zisserman A., Fitzgibbon A., 1999: Automatic line matching and 3D reconstruction of buildings from multiple views. IAPRS, Munich, Germany, Vol. 32, Part 3-2W5, pp. 69-80

Braun A., 1999: Automatic detection and 2-D reconstruction of buildings from orthoimages using vectorized map information. Diploma thesis, Institute of Geodesy and Photogrammetry, ETH Zürich

Eidenbenz Ch., Käser Ch., Baltsavias E., 2000: ATOMI - Automated reconstruction of Topographic Objects from aerial images using vectorized Map Information. IAPRS, Amsterdam, Netherlands, Vol. XXXII

Gruen A., Kübler O., Agouris P. (Eds.), 1995: Automatic Extraction of Man-Made Objects from Aerial and Space Images. Birkhäuser Verlag, Basel

Gruen A., Baltsavias E., Henricsson O. (Eds.), 1997: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser Verlag, Basel

Haala N., Walter V., 1999: Automatic classification of urban environments for database revision using lidar and color aerial imagery. IAPRS, Valladolid, Spain, Vol. 32, Part 7-4-3 W6, pp. 76-82

Henricsson O., 1996: Analysis of image structures using color attributes and similarity relations. Dissertation, Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland

Maas H.-G., 1999: Closed solutions for the determination of parametric building models from invariant moments of airborne laserscanner data. IAPRS, Munich, Germany, Vol. 32, Part 3-2W5, pp. 193-199

Streilein A., 1999: Luftbildgestützte automatische Generierung von 3D Gebäudeinformationen aus topographischen Karten. Proceedings of the 10th international geodetic week 1999, Obergurgl, Austria, pp. 183-192

Sibiryakov A., 1996: House detection from aerial color images. Internal Report, Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland

Weidner U., 1997: Digital Surface Models for Building Extraction. Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag, Basel, pp. 193-202

Pratt W. K., 1991: Digital image processing. John Wiley & Sons Inc., New York

Zhang Ch., Baltsavias E., 2000: Image analysis for 3-D edge extraction and knowledge-based road reconstruction. IAPRS, Amsterdam, Netherlands, Vol. XXXII