3-D building reconstruction with ARUBA: a qualitative and quantitative evaluation

Conference Paper

Author(s):
Henricsson, Olof; Baltsavias, Emmanuel P.

Publication date:
1997

Permanent link:
https://doi.org/10.3929/ethz-a-004334529

Originally published in:
https://doi.org/10.1007/978-3-0348-8906-3_7

3-D Building Reconstruction with ARUBA: A Qualitative and Quantitative Evaluation

Olof Henricsson and Emmanuel Baltsavias

Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology
ETH – Hönggerberg, CH-8093 Zürich, Switzerland
e-mail: {ohenri,manos}@geod.ethz.ch

Abstract

Reliable and accurate 3-D reconstruction of man-made objects is essential for many applications using digital 3-D city models. Manual reconstruction of buildings from aerial images is time consuming and requires skilled personnel, hence large efforts are being directed towards the automation of building detection and reconstruction. In this paper we present ARUBA[1] — a framework for automated 3-D building reconstruction. After highlighting our strategy and concisely describing the framework and its modules, we evaluate the reconstructed roofs relative to accurate reference data based on three criteria: completeness, geometric accuracy and shape similarity. Finally, we interpret the results of the performance evaluation and make suggestions for improvements.

1 Introduction

Analysis of digital aerial images has been an active research topic in the Computer Vision community as well as in Photogrammetry for a number of years. Automated methods for reliable and accurate 3-D reconstruction of man-made objects are essential to many users and providers of 3-D city data, including urban planners, architects, and telecommunication and environmental engineers. A 3-D city model captures primarily the geometric description of all objects of interest in an urban area in computer compatible form. Objects of interest include for example buildings, terrain, vegetation, traffic networks, and public utilities. This paper deals with the most important urban object — buildings. Manual 3-D processing of aerial images is time consuming and requires the expertise of qualified personnel and often expensive equipment. Therefore, the necessity to interpret and quantitatively process digital aerial images in a semi- or fully automatic mode using a standard computer and to integrate the results into CAD- or spatial information systems is more urgent than ever.

[1] ARUBA: Automatic Reconstruction of Buildings from Aerial Images

Early work on building extraction/detection based their processing on single grey-valued images applying heuristics, simple object models, and shadow analysis to solve the *building detection* problem. The main task was not to reconstruct the building in 3-D but to detect it and find its 2-D outline. These approaches rely on the assumption that man-made objects possess a large amount of geometric regularity (e.g. flat rectilinear roofs), which is explicitly used to reduce the number of building hypotheses. However, even when relying on such simple roof models the analysis of monocular images is an extremely difficult task since it generally leads to ambiguous solutions.

In the last few years, several (academic) groups have presented new promising results in automated 3-D building reconstruction, for example (Roux and McKeown 1994, Haala and Hahn 1995, Lang and Förstner 1996, Weidner 1996, Wiman and Axelsson 1996, Noronha and Nevatia 1996). Shadow analysis, as the main cue for inferring 3-D structures from monocular images, has been abandoned for 3-D processing techniques using 3-D information such as a Digital Surface Model (DSM), 3-D edges, and 3-D corners extracted from multiple, overlapping aerial images. With this aerial imagery, building roofs can be *reconstructed in 3-D*. Vertical walls may be added afterwards by projecting the eaves of the roof down to an existing Digital Terrain Model (DTM).

With this paper we demonstrate that it is possible to automatically reconstruct the roof of buildings even when the shapes of the 3-D parts is not known a priori. ARUBA is a general framework for automatic reconstruction of building roofs from high resolution aerial images (Henricsson et al. 1996, Henricsson 1996, Bignone et al. 1996) and is designed to reconstruct a general class of roof types with a high metric accuracy. The strategy employed consists in extracting planar 3-D patches, which are then assembled to complete roofs. A generic 3-D patch is non-vertical, roughly planar and encloses a compact polygonal 2-D enclosure with similar photometric and chromatic attributes along its boundary. ARUBA relies on hierarchical hypothesis generation in both 2-D and 3-D, thereby using procedures for feature extraction, segment stereo matching, 2-D and 3-D grouping, and color and object modeling. We argue that geometric regularity, although important, cannot serve as sole basis for extracting complex structures for which no generic models exist.

The main objectives with this paper are twofold: to present the building reconstruction framework and to evaluate the reconstructed roofs of buildings with respect to accurate reference data. Section 2 describes our strategy for automated 3-D building reconstruction. After a short presentation of the framework (section 3), we evaluate the reconstructed building roofs with respect to accurate reference data (section 4). We focus on three criteria: completeness, geometric accuracy and shape similarity. We interpret the results of the assessment and make an in-depth analysis of the underlying causes. Based on this analysis we make suggestions for improvements (section 5).

2 A Strategy for 3-D Building Reconstruction

The goal of ARUBA is to automatically reconstruct the main 3-D roof structure of a general class of buildings with a high metric accuracy using high resolution aerial images. Based on this objective we describe the main features of our strategy for 3-D building reconstruction.

1. make use of multiple, overlapping color images
2. early transition to 3-D data (3-D contours and planes)
3. generic object modeling directly in 3-D (set of adjoining 3-D patches)
4. make use of general object knowledge (geometric, surface)
5. mutual interaction between 2-D and 3-D processes
6. separation of building detection and reconstruction

Color images indisputably contain more information than grey-valued images. The main issue is how this color information can be used in generic object reconstruction. In this work, we generally assume that each roof surface is locally perceptually uniform along its boundary. This does require that the surface has a particular color, e.g., red, it just states that boundary of a roof is assumed to have locally similar spectral characteristics. Color cues, in the form of color region attributes, are used together with geometric cues to considerably improve the performance of 2-D and 3-D grouping, hence making it possible to reconstruct also complicated roof shapes.

We propose an approach to 3-D building reconstruction which consists in extracting generic planar patches, which are then assembled to complete roofs. A generic 3-D patch is non-vertical, roughly planar and encloses a compact polygonal 2-D enclosure with similar photometric and chromatic attributes along its boundary. By modeling not only the geometry of the roof, but also spectral properties along its boundary we can handle a large variety of roof shapes. General knowledge about the geometry of roof parts (e.g. boundary length and shape complexity) and surface characteristics (e.g. color homogeneity) are thereby be used to reduce the algorithmic complexity in generating hypotheses.

Whenever 3-D features are incomplete or entirely missing, additional (more complete) 2-D information can be used to infer the missing features and structures. This further means that a mutual interaction between 2-D and 3-D procedures is required at certain levels of processing (see Fig. 1). This interaction, which considerably reduces the search space and thereby also the overall complexity, is important since neither 2-D nor 3-D procedures *alone* are sufficient to solve the problems.

In order to further reduce the complexity of the reconstruction task, we assume that each building is presented in isolation, i.e., the detection of the buildings is already done. Building detection can either be automatic or manual, depending on the complexity of the scene and its main objective is to generate regions of interest enclosing the same building in all images. Notice that, this does not imply that buildings are isolated. We use the operator to mark a window enclosing the same building in all overlapping images, see Fig. 1. After this initialization, the building is *automatically reconstructed*.

3 Automatic Reconstruction of Buildings (ARUBA)

The ARUBA system employs a simple but very powerful modeling approach – a complete roof consists of a set of planar parts which mutually adjoin along their boundary. Because of this requirement the framework can not reconstruct roof parts that adjoin another roof *inside* its boundary, see the dormer windows in Fig. 1.

The ARUBA framework is shown in Fig. 1, with its 2-D processing modules located in the light grey area. The first processing step involves extracting a contour graph, including edges, lines and key-points. As the contour graph contains only basic information about geometry and connectivity we increase its usefulness by assigning rich attributes to each contour. The attributes reflect either properties of the contour (e.g. length, integrated gradient magnitude, edge/line type) or region properties on either side (e.g. photometric and chromatic). The photometric and chromatic region attributes are computed for each contour by finding color clusters using the CIELAB color space and robust estimation procedures. For details on these issues, we refer to (Henricsson 1996).

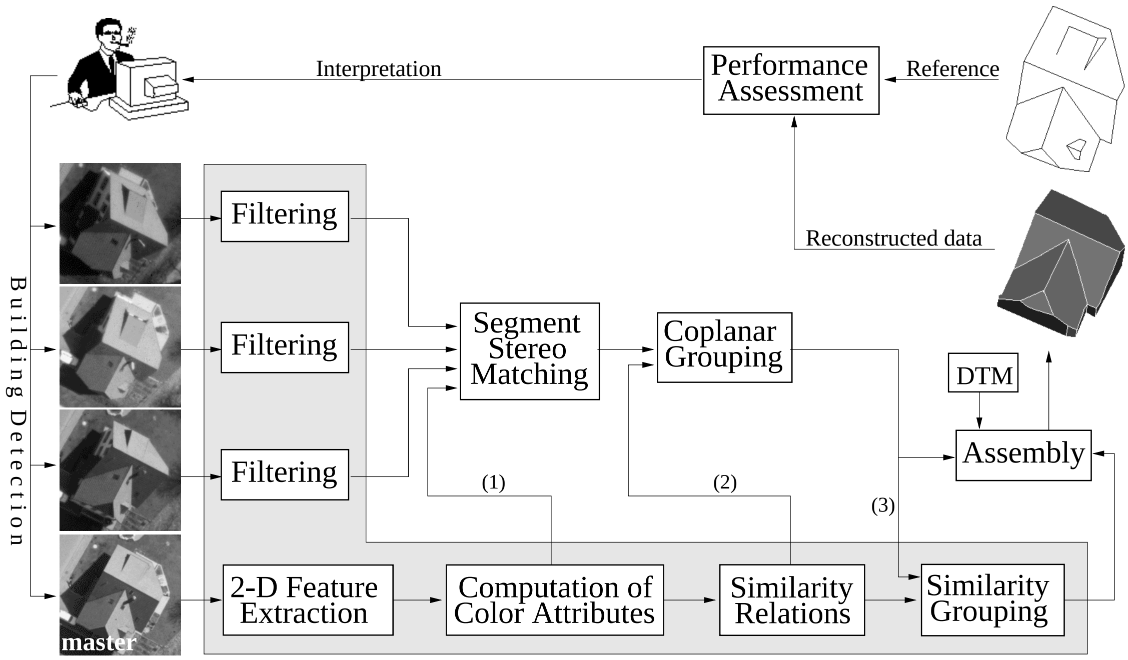


Fig. 1: The ARUBA system for fully automatic 3-D reconstruction of buildings. The operator detects the buildings by marking a rectangular window enclosing the same building in all four images. The subsequent reconstruction in 3-D is fully automatic. The 2-D processing modules are located in the light grey area.

Straight 2-D contours from a master image are matched in the other images using a novel approach to stereo matching. Edges are extracted from *only one* image (the master image) and are matched in the other images by maximizing an “edginess measure” along the epipolar line. The edginess measure is a function of the gradient (magnitude and direction) in the other images. Geometric and photometric constraints are also used to reduce the number of mismatches. The algorithm produces a set of 3-D segments. These 3-D segments are then grouped to hypotheses of planes by means of coplanar grouping. Both these algorithms are described in (Bignone 1995, Bignone et al. 1996). In most cases, only a subset of all 3-D segments on each plane actually represents the outer boundary of a roof. Furthermore, the planes are often incomplete due to false matches or when the matching algorithm does not find the correct correspondences for the 2-D contours. The extracted planes themselves are therefore not sufficient to describe the roofs. The object boundary of each plane hypothesis is found by extracting 2-D enclosures employing a novel grouping technique, *similarity grouping*, which is based on similarity in proximity, orientation, and photometric and chromatic region attributes (Henricsson 1996). The extracted 3-D contours are used as "seed structures" thereby reducing the complexity of grouping. Finally, the most evident and consistent set of planar roof hypotheses is selected based on simple geometric criteria in 2-D and 3-D. Vertical walls are added afterwards by projecting the eaves of the roof down to a DTM. The end result is a complete 3-D model of the roof and its vertical walls, including 2-D contours, 3-D segments, 3-D planes and their topology.

The interaction between 2-D and 3-D processing is important. In most other works this interaction is restricted to a merging step of 2-D and 3-D features and structures without having mutually exchanged information during the processing. In Fig. 1 it essentially means to remove the three marked interactions. The shortcoming of such a interaction becomes obvious when dealing with more complex objects than the usual ones. In our approach we exploit information from other modules to reduce the complexity at each processing level, thereby also reducing the overall complexity. We venture to say that the design of the algorithmic framework, including its flow and interactions, is at least as important as developing high quality processing modules.

The main difference of our approach with those of other groups consists in the extensive use of color attributes and similarity relations combined with the overall aim to reconstruct a general class of buildings. The importance of using color cues in building reconstruction manifests in the fact that all subsequent processing builds upon this information. Color cues, in the form of color region attributes, are a prerequisite for similarity grouping and thus for generating hypotheses of generic roof parts. Also the 3-D processes, i.e., stereo matching and coplanar grouping, strongly benefit from exploiting this data.

4 Performance Assessment

Performance assessment involves evaluating the performance of single modules and the complete system, as well as the qualitative and quantitative assessment of the results. Because ARUBA is still an experimental framework we do neither address the performance of single modules nor the complete system (e.g. computation times, memory usage, sensitivity of the parameters). Instead we assess the completeness, geometric accuracy and shape similarity of the reconstructed roofs relative to accurate reference data.

4.1 Evaluation of the Results

As test region, we choose the residential scene from the Avenches data set (Mason et al. 1994), mainly due to the availability of multiple, overlapping, color images and accurate reference data. This high precision photogrammetric data set has the following characteristics: 1:5,000 image scale, vertical aerial photography, four-way image coverage (60% forward and side-wards overlap), flying height approx. $750\ m$, color imagery of size 1800×1800 pixels, geometrically accurate film scanning with 15 microns pixel size (i.e. ground area of approximately 7.5×7.5 cm), precise sensor orientation, and accurate ground truth including a DTM and buildings. The reference data of the buildings was manually acquired by an experienced operator at an analytical plotter to an estimated accuracy of ± 10 cm.

Figure 2 show in (A) one of four overlapping color aerial images and (B) the corresponding ARUBA reconstruction, where twelve of thirteen roofs have been successfully reconstructed. The algorithms fail to reconstruct building no. 4, which is under construction (covered with blue plastic sheets). Building no. 9 is very complicated because a group of trees cast large shadows on the right roof part in all four views. As a consequence, the right roof patch is not homogeneous enough to allow a correct reconstruction (i.e., only the left roof part is reconstructed). To reconstruct the entire roof (as in Fig. 2), the lightness and color homogeneity criteria had to be modified.

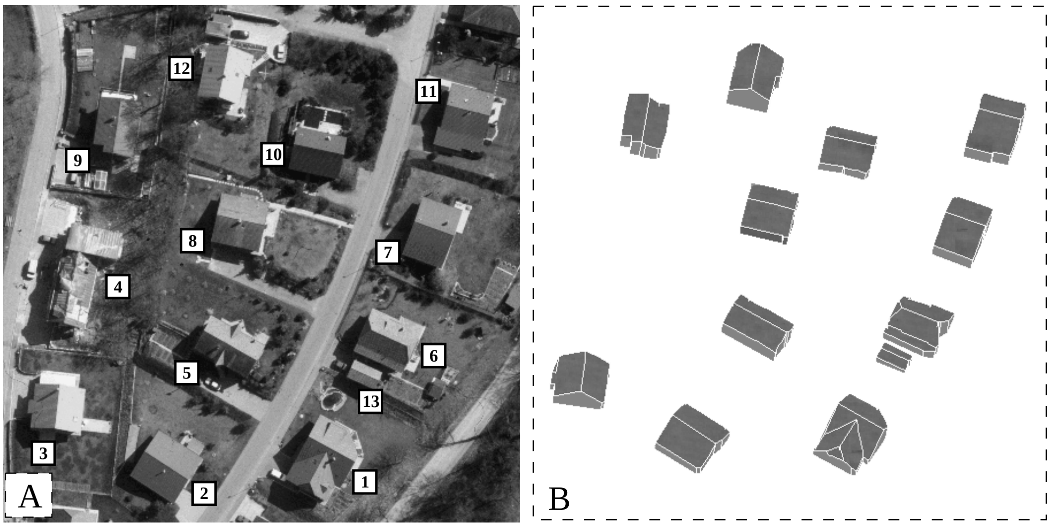


Fig. 2: The ARUBA building reconstruction. **(A)** one of four overlapping color aerial images, and **(B)** one view of the 3-D building reconstruction.

The assessment of the reconstructed buildings with respect to the reference data is performed by comparing corresponding roof parts. The assessment of the entire building is then the combination of all part errors. Discrepancies between the reconstructed roofs and the reference data can be subdivided into: missing or additional roof parts, rotation and translation differences between the planes, and differences in area and shape. Here, we are mainly interested in the total error of the reconstruction. We have therefore defined three different measures: completeness, geometric accuracy, and shape similarity. The first measure, completeness, refers to the number of reconstructed parts with respect to the reference. Geometric accuracy is divided into the displacement in the normal direction between the two planes and their difference in orientation. Shape similarity is very difficult to quantify, therefore, we chose to use two weak dissimilarity measures: the difference in area and overlap error, both computed with respect to the reference area.

To assess the geometric accuracy of the reconstructed roofs we start by analyzing individual roof parts. We fit a plane to the reference coordinates of each roof part and then project them onto the fitted plane. Apart from an improvement vector for each reference coordinate, the fit procedure also returns the overall rms-values in lateral and vertical directions and the normal vector of the plane. The same fit procedure is also applied to the reconstructed data.

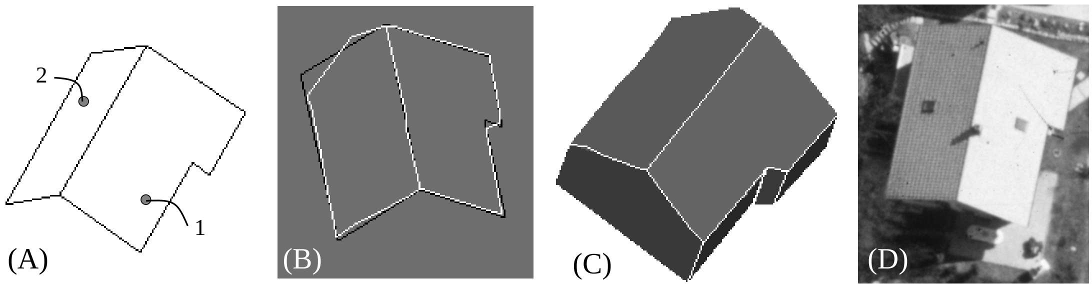


Fig. 3: Building no.12: **(A)** the reference data, **(B)** the reference data (black) overlaid with the reconstructed roof (white), **(C)** the reconstructed building including the roof and vertical walls, **(D)** the original "color" image. Notice the poor contrast in the upper left corner.

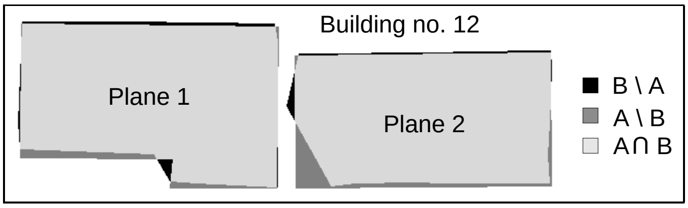
The difference in plane orientation and the displacement between the two planes are two important accuracy measures. In Table 1, we have listed the difference in orientation, the absolute normal $|d_n|$ and vertical d_z displacements between the reference and the reconstructed planes for each roof part of building no. 12 (see also Fig. 3). The negative sign of d_z means that the reconstructed roof is above the reference.

	Reference Plane			**Reconstr. Plane**			**Difference of Planes**		
	RMS $[cm]$			**RMS** $[cm]$			**Angle**	**Distance** $[cm]$	
Plane	x	y	z	x	y	z	$[deg]$	$\|d_n\|$	d_z
1	2.67	1.70	4.09	0.96	0.61	1.48	1.0	6.2	-4.9
2	1.55	0.92	2.25	0.67	0.40	0.99	0.4	17.3	-13.7
						$mean =$	**0.7**	**11.7**	**-9.3**

Tab. 1: Building no. 12. The table lists rms-values for both the reference and the reconstructed planes, the difference in orientation, and the displacement in normal and vertical directions. The mean (total) values for the entire roof are also computed.

Notice in Table 1, that the rms-values for the reconstructed roof parts are significantly smaller than those of the reference data. In fact, this is true for all evaluated buildings. We found that the rms is often below 2-3 centimeters in planimetry which is two to three times smaller than corresponding values in the vertical direction. Further, the rms of the reconstructed planes are often considerably smaller than those of the reference planes. This fact is not surprising, since the reconstructed roof part is the result of a coplanar grouping procedure, which uses only 3-D segments fulfilling weak co-planarity constraints.

Although results are excellent, the displacement and orientation values in Table 1 only cover certain aspects of the geometric accuracy of the reconstruction and nothing about the similarity in shape and area. To compute shape dissimilarity, we first project the fitted coordinates of the reconstructed roof part onto the reference plane without performing any rotation or translation. This means that rotation and translation errors are also included. Each 3-D patch is then transferred to a 2-D pixel map using a $5 \times 5\ cm$ grid size. The area of each of the reference (A) and the reconstructed (B) roof parts are approximated by the number of pixels. The figure in Table 2 shows the overlapping planes for building no 12. The intersection area of the two planes $A \cap B$ is labeled in light grey. The set difference between the reference and the reconstructed planes (and vice versa) is defined as $A \setminus B$ ($B \setminus A$) and illustrated in the example.



Plane	Area in $[m^2]$						$\frac{\lvert A-B \rvert}{A}$	$\frac{A \setminus B + B \setminus A}{A}$
	A	B	$\lvert A-B \rvert$	$A \cap B$	$A \setminus B$	$B \setminus A$	[%]	[%]
1	92.41	91.54	0.87	88.93	3.48	2.61	0.9	6.6
2	84.14	80.12	4.02	78.49	5.66	1.64	4.8	8.7
Total	**176.55**	171.66	**4.89**	167.42	**9.14**	**4.25**	**2.8**	**7.6**

Tab. 2: Data for building no. 12. The table lists the area for both the reference (A) and the reconstructed (B) planes, the arithmetic difference $|A - B|$, intersection $A \cap B$, set differences $A \setminus B$ and $B \setminus A$, and the two shape similarity ratios between the reference (A) and the reconstructed (B) roof parts.

The intersecting area $A \cap B$ with respect to the reference area A is not a good indicator of the quality of the results, since it does not account for cases where A is completely included in B and $B \gg A$. Instead, we choose to work with two relative error measures: the relative arithmetic difference $\frac{|A-B|}{A}$ and the sum of the set differences divided by the reference area $\frac{A \setminus B + B \setminus A}{A}$. These two ratios should both be small. The latter is the sum of the area difference and the remaining overlap error, i.e., the *total* relative shape dissimilarity. Table 2 lists these measures and values for building no. 12 and Table 3 the corresponding data for all buildings. Notice in Table 3, that the computations were performed for all planes and buildings that have been successfully reconstructed.

4.2 Interpretation and Discussion

The reference data (excl. building no. 4) includes 12 buildings with 39 planes. The ARUBA software reconstructed 29 planes. No extra roof parts have been extracted, however some are missing. Among the ten missing planes, only one of them is part of a main roof structure, i.e. one of the two triangular planes is missing for building no. 6 due to vanishing contrast. The remaining non-reconstructed planes all belong to smaller structures such as dormer windows (see for example house no. 1 and 5). Remember that the reconstruction algorithm was deliberately not designed to handle planes that do not adjoin along the boundary. If these smaller roof structures can be properly extracted, we may add them in a second assembly stage. The area of the 29 reconstructed planes constitutes 97.5% of the total reference area, which confirms that the missing structures are no essential components of the roof (although the people living there might not agree).

Bldg	Rec	Angle	$\bar{\lvert d_n \rvert}$	$\bar{d_z}$	**Area in** $[m^2]$				$\frac{\lvert A-B \rvert}{A}$	$\frac{A \setminus B + B \setminus A}{A}$
No	Plns	$[deg]$	$[cm]$	$[cm]$	A	$\lvert A-B \rvert$	$A \setminus B$	$B \setminus A$	[%]	[%]
1	5/9	1.7	24.2	-15.0	281.56	10.29	16.17	15.07	3.7	11.1
2	2/3	2.4	24.6	-21.3	201.09	3.41	6.18	7.32	1.7	6.7
3	2/2	1.2	7.4	-5.8	179.01	1.05	7.10	6.04	0.6	7.3
5	2/6	2.2	4.2	3.4	165.71	6.29	1.39	7.67	3.8	5.5
6	4/5	4.5	16.0	-8.8	211.00	5.14	11.90	8.49	2.4	9.7
7	2/2	1.0	15.9	-13.8	178.00	2.51	1.82	4.33	0.6	3.5
8	2/2	1.2	10.8	-9.7	169.38	4.24	6.93	2.71	2.5	5.7
9	2/2	1.0	10.7	-10.1	145.90	6.35	11.17	4.81	4.4	11.0
10	2/2	2.0	9.2	-7.2	166.00	3.05	3.00	6.06	1.8	5.5
11	2/2	1.4	14.8	-13.6	183.84	3.01	4.92	3.00	1.7	4.3
12	2/2	0.7	11.7	-9.3	176.55	4.89	9.14	4.25	2.8	7.6
13	2/2	5.4	9.5	-5.0	38.06	2.99	1.58	3.89	7.9	14.4
Total	**29/39**	**2.2**	**14.6**	**-10.2**	2096.10	53.22	81.30	73.64	**2.5**	**7.4**

Tab. 3: Evaluation data for all reconstructed buildings in the Avenches data set. The second column lists the completeness values for each building, whereas the following three columns show the geometric accuracy of the reconstruction: the difference in orientation and the displacement in normal and vertical directions between the reference (A) and reconstructed (B) planes. The six rightmost columns show the total roof area of the reference, the arithmetic and set differences, and the two relative shape dissimilarity measures.

The difference in orientation between the reference and the reconstructed planes is an important accuracy measure together with the displacement. The bottom row in Table 3 reveals that the average difference in orientation is 2.2 degrees, which we believe is a good result. Two buildings (no. 6 and 13) have larger orientation differences. For both these buildings, no single plane is way off the average, which means that these two buildings are generally less accurately reconstructed. These errors come from the segment stereo matching and coplanar grouping procedures.

The average *absolute* displacement $|d_n|$ between the reference and reconstructed planes for all buildings is 14.6 centimeters. The relation between the estimated average displacement error and the flying height is $0.19‰$, which is a respectable result also for manually measured objects. The average displacement (with sign) in the vertical direction d_z is -10.2 centimeters, which means that the reconstructed planes are on average one decimeter above the corresponding reference planes in the vertical direction. This vertical shift has been noticed through visual inspection, however until now we have not been able to verify it. This systematic shift comes from differences in orientation parameters. For the generation of the reference data a stereo model orientation on an analytical plotter was used while for the reconstructed data the orientation comes from a bundle adjustment of the whole image block.

In general, the perceptual quality of the results is impressing. All reconstructed buildings actually look like buildings and their shape is close to the reference data, i.e., the human interpretation. The total relative shape dissimilarity, $\frac{A \setminus B + B \setminus A}{A}$, reflects some aspects of the quality of the reconstructed roof parts. However, the rotation and translation effects are also included, which in some cases is the dominating error source. The ratio $\frac{|A-B|}{A}$ represents the difference in area between the two planes relative to the reference area and comes from missing or protruded parts along the boundary or through rotation effects. Except for building no. 13 this error is small — on average 2.5%. This error is included in the total relative shape dissimilarity $\frac{A \setminus B + B \setminus A}{A}$, which is the combination of two errors: the difference in shape (form) and a translation or a rotation between the two planes.

Analyzing the results in Fig. 3 and Table 2, we first notice that both planar patches have been successfully reconstructed, plane 1 qualitatively better than plane 2. In both cases the reference area is larger than the reconstructed, with a larger area difference for plane 2. Analyzing the ratios in the two rightmost columns in Table 2 we see that, for plane 2, the main error source is the difference in area, which comes from the missing part of the boundary (a corner is cut off, see Fig. 3D). The dominating overlap error for plane 1 originates in the differences in rotation and translation between the planes.

The rightmost column in Table 3 list the total relative shape dissimilarity measures for each building. We notice that building no. 13 (garage) shows the poorest results. The extracted 3-D contours are poorly located and hence also the 3-D planes. In addition, several reference coordinates are poorly measured which adds to the effect. The most complicated buildings, i.e., no. 1 and 6, show slightly poorer results than the simpler ones. We also notice the large shape dissimilarity for building no. 9. This high value comes primarily from the poorly reconstructed right roof part. In general, we believe that a good definition of the roof boundary is more important than the complexity of the shape, which would indicate that the ARUBA framework can deal with different roof types in an equal manner.

The major single error source is the segment stereo matching (Bignone 1995). The stereo matching influences primarily the *completeness and geometric accuracy*. However, if many 3-D segments are missing and hence also important 3-D planes, then we cannot expect a successful reconstruction of the building. Therefore, a robust and accurate stereo matching procedure is crucial for a high quality reconstruction. Assuming that the 3-D segments and planes are adequately extracted, then the *shape similarity* of the reconstructed planes with respect to the ground truth depends mostly on the similarity grouping and on the selection criteria in the assembly procedure, see Fig. 1.

To summarize the evaluation of the reconstruction results we venture to say that the fully automatic ARUBA reconstruction produces comparable results to those of manual measurements. The accuracy and completeness of the ARUBA reconstruction lies in the range of the expected accuracy for the reference data. The shapes of the reconstructed roof parts are similar to those of the reference data, which were interpreted online by a human operator during data acquisition. The total relative shape dissimilarity is an adequate indicator of the shape quality of the reconstructed buildings, even though the measure also includes the rotation and translation errors.

5 Outlook

The general design of ARUBA is conceptually sound and the algorithms produce good results, both qualitatively and quantitatively, in suburban areas. However, the algorithms are not capable of fully automatically handling certain complex scenes in suburban and densely populated urban regions in general due to connected buildings, shadows from trees and other buildings, and occlusion situations. Apart from improving certain processing modules in the existing framework (e.g. segment stereo matching), we see two important extensions: modeling of roof parts and user/machine interface.

One important extension to the ARUBA framework involves object modeling, i.e., modeling of roof parts and of complete roofs. In our opinion, the main deficiency of the ARUBA system is the modeling aspect. Exploiting that most *roof parts* have a simple shape should allow the system to more effectively handle the generation of 3-D patches. A verification (or self-diagnosis) also require object models. Generic models of roof parts may include geometric shapes such as rectangular, triangular, parallelograms and a few other primitive shapes occuring in abundance, as well as surface characteristics.

Sound concepts for human/machine interface and feedback are crucial for a successful reconstruction especially in urban areas, since we do not expect the automatic system to correctly handle all scenes. The human operator may be involved in initialization/detection, providing model information, imposing constraints (e.g. rectilinear), and marking/editing incorrect reconstructions.

6 Conclusions

We have presented ARUBA, an experimental framework for automated 3-D reconstruction of buildings from aerial images. The approach makes effective use of much of the available 2-D and 3-D information present in the images of a given site. Geometry, photometric and chromatic attributes and stereo information about contours and their flanking regions are effectively combined. Consequently, the procedure is more robust than one that uses only partial information. This approach has proven powerful enough so that, in contrast to most approaches to building reconstruction, we need not assume the roofs to be flat or rectilinear or use a parameterized model of the complete building.

We have evaluated the reconstruction results on the Avenches data set with respect to accurate reference data. We focussed on three criteria: completeness, geometric accuracy and shape dissimilarity. We have shown that the average difference in orientation is $2.2\ [deg]$ and that the average absolute displacement between the planes is $14.6\ [cm]$. The displacement in vertical direction is $-10.2\ [cm]$, which indicates a systematic error (probably difference in orientation parameters). We further developed a relative measure for shape dissimilarity between the reference and the reconstructed roofs. The most complicated buildings show slightly larger shape dissimilarity than the simpler ones, which is in line with the interpretation of a human operator.

To conclude, the fully automatic ARUBA reconstruction produces comparable qualitative results to those of manual measurements and the geometric accuracy lies in the range of the expected accuracy for the reference data.

Acknowledgments

The authors cordially thank Gaudenz Danuser for invaluable support in developing the evaluation procedures and also in the interpretation of the results.

References

Bignone F. (1995) *Segment Stereo Matching and Coplanar Grouping*, Technical Report BIWI-TR-165, Institute of Communications Technology, Image Science Lab, ETH, Zürich, Switzerland.

Bignone F., O. Henricsson, P. Fua, M. Stricker (1996) *Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery*, in B. Buxton, R. Cipolla (eds), Computer Vision – ECCV'96, Springer Verlag, Berlin, pp. 85–96.

Haala N., M. Hahn (1995) *Data fusion for the detection and reconstruction of buildings*, in A. Grün, O. Kübler, P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag, Basel, pp. 211–220.

Henricsson O. (1996) *Analysis of Image Structures using Color Attributes and Similarity Relations*, PhD thesis, Swiss Federal Institute of Technology (ETH) Zurich No. 11663, published in *Mitteilungen Nr. 59* at the Institute of Geodesy and Photogrammetry.

Henricsson O., F. Bignone, W. Willuhn, F. Ade, O. Kübler, E. Baltsavias, S. Mason, A. Grün (1996) *Project Amobe, Strategies, Current Status, and Future Work*, International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, pp. 321–330.

Lang F., W. Förstner (1996) *Surface Reconstruction of Man-Made Objects using Polymorphic Mid-Level Features and Generic Scene Knowledge*, International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, pp. 415–420.

Mason S., E. Baltsavias, D. Stallmann (1994) High Precision Photogrammetric Data Set for Building Reconstruction and Terrain Modelling, *Internal Report, Institute of Photogrammetry and Geodesy*.

Noronha S., R. Nevatia (1996) *Detection and Description of Buildings from Multiple Aerial Images*, DARPA Image Understanding Workshop, Palm Springs, CA, pp. 469–478.

Roux M., D. McKeown (1994) *Feature Matching for Building Extraction from Multiple Views*, DARPA Image Understanding Workshop, Monterey, CA, pp. 331–349.

Weidner U. (1996) *An Approach to Building Extraction from Digital Surface Models*, International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, pp. 924–929.

Wiman H., P. Axelsson (1996) *Finding 3D-Structures in Multiple Aerial Images Using Lines and Regions*, International Archives of Photogrammetry and Remote Sensing, Vol. XXXI, Part B3, pp. 953–959.