3D CITY MODELLING FOR MOBILE AUGMENTED REALITY

Siyka Zlatanova, Frank A. van den Heuvel
Delft University of Technology, Department of Geodesy
Thijsseweg 11, 2629JA Delft, The Netherlands
Email: {S.Zlatanova, F.A.vandenHeuvel}@geo.tudelft.nl

KEYWORDS: city modelling, close-range, aerial photogrammetry, laser altimetry, data integration, augmented reality

ABSTRACT

The outdoor augmented reality is one of the relatively new applications that allows the user to observe virtual scenes superimposed onto the real view. A key issue in such augmenting of the real world is the precise alignment of virtual and real objects that is closely related to the positioning and orientation of the user. This paper presents an approach for 3D reconstruction and creation of a 3D model that is intended for an augmented reality system. The paper discusses the role of the 3D model in the augmented reality system and clarifies representation, resolution and accuracy requirements. The 3D objects are organised in a typical 3D topological data structure extended to comprise line features. The presented reconstruction methods are separated into two major groups, i.e. methods for reconstructing buildings and methods for reconstructing terrain objects (streets, parking lots, etc.). The reconstruction of buildings is based on manual measurements on terrestrial and aerial images. Depending on the complexity and importance of buildings, further distinction in the procedures is made for complex and simple buildings. The complex buildings are reconstructed manually, utilising commercial software. A semi-automatic procedure consisting of manual digitising of roof outlines from aerial images, and automatic assembling of walls is used for reconstructing relatively simple buildings. Terrain objects are obtained automatically from a large scale digital topographic map and laser altimetry data. The utilised methods and the accuracy of the obtained results are deliberated in detail. The paper concludes with a discussion on practical problems related to the 3D reconstruction and integration of data in the 3D topological data structure.

1 INTRODUCTION

3D reconstruction of urban environments is a topic of intensive research in the last several decades due to increasing demand for 3D models. A number of automatic and semi-automatic approaches and systems for 3D reconstruction of man-made objects (e.g. buildings) have been progressing rapidly. However, many problems still remain unsolved since no approach is suited for all the applications, types of objects, resolution and accuracy requirements. Augmented reality is one of the applications where 3D models of large urban areas or individual buildings are becoming an important factor. Augmented reality is a combination of a real scene seen by the user and a virtual scene generated by the computer. The virtual scene augments the real scenes with additional information, e.g. newly projected buildings, underground pipe lines, text and symbols for orientation. A device known as Head Mounded Display's (HMD) allows the user to observe simultaneously virtual and real views. Such augmenting of the real world can facilitate and ease a number of urban activities, e.g. urban planning, utility maintenance, public rescue operations, etc. One of the key issues in the development of a successful application is the accurate alignment of virtual and real objects that is closely related to the problems of 3D modelling.

This paper presents our integrated approach to 3D reconstruction of urban models for the purpose of mobile augmented reality. The research is a part of the interdisciplinary UbiCom project that is conducted at the Delft University of Technology, The Netherlands (UbiCom 2001).

2 THE ROLE OF THE 3D MODEL IN UBICOM

The importance of the 3D modelling can be outlined better by giving a short explanation of the processes in which the 3D model is involved. Figure 1 gives a broad idea of the entire system. A question of the user (e.g., "which is this building?", "where is the library?", "what the new statue at front of the Aula looks like?") is sent to the receiving stations and the backbone computers for processing. The answer in form of text or graphics is transmitted back to the user and visualised in the HMD. Depending on the question, the user may receive: text information visualised near the building of interest, or 2D virtual arrows pointing the way to the library, or 3D model of the statue placed at the project location. In either case, the accurate alignment of virtual and real views is critical for the human perception. Small movements of the user may cause significant displacement of the virtual objects that may further mislead the user. The role of the 3D model in UbiCom is related exactly to the accurate alignment of virtual and real objects.

Two subsystems of the augmented architecture utilise the 3D model for two different purposes. First, the 3D model is to be used for the accurate positioning of the user in the real world. In principle, the pose determination in the UbiCom system is based on a vision system. The user is equipped with a mobile set consisting of a video camera, GPS and an inertial system. GPS and the inertial system are expected to give an initial positioning with accuracy approximately two meters (Persa and Jonker 2001). This accuracy is not sufficient for aligning virtual and real objects observed at close distances. For example, if an urban planner would like to justify the position of a statue in front of a building, he/she might need to have it visualised in the HMD with an accuracy of at least 10 cm. The accurate positioning is to be achieved with the help of the 3D model, i.e. by a matching procedure between features extracted from the video images (obtained from the mobile video camera) and features retrieved from the 3D model. Within the UbiCom project, the mobile equipment is intended as a light unit with limited power supply, which imposes limitations on the features for matching. Currently, we have concentrated on line matching. This means that the 3D model has to provide appropriate line features, i.e. well distinguished lines on the facades of buildings, on the ground (e.g. tiles, bicycle paths) or on other objects (e.g. lampposts, traffic signs) in the field of view. Furthermore, the number of lines supplied by the 3D model plays a critical role in the matching procedure and hence for the accurate positioning.

Second, the rendering subsystem (for visualising virtual objects) makes use of the 3D model but in a different way. In order to achieve realistic visualisation, the virtual objects have to "behave" as real objects, i.e. when occluded by real objects, the corresponding parts have to be invisible for the observer (Pasman and Jansen 2001). To achieve this effect, the rendering engine must know the exact position of the mobile unit and the position and the shape of the occluding real object. Currently, we concentrate on large 3D man-made objects (e.g. buildings, monuments, bridges) as potential occluders. Real objects such as trees, cars, windows, doors, balconies, etc. are not considered. The geometric representation of potential occluders has to assure connectivity and continuity, i.e. gaps between polygons or polygons with holes are not acceptable since they may disturb the perception of the mixed scene. In other words, 3D topological consistency is required.

The analysis of the role of the 3D model in the augmented reality system can be summarised into three basic requirements as follows:

- maintenance of topologically structured 3D objects,
- accuracy in the range of few decimetres, and
- large amounts of details on facades organised as individual line features.

In this paper we concentrate mainly on the procedures for reconstructing topologically structured 3D objects with decimetre accuracy.

The test area for the outdoor augment reality system is the central campus area of the Delft University of Technology (see Figure 2). Although the area is relatively small, the 3D objects of interest revealed large variations in shape and complexity. Here, we concentrate on the reconstruction of five 3D objects (see Figure 2) namely the Aula (1), the Faculty of Applied Physics (2), the Faculty of Mechanics (3), the Post-office (4), and the Art monument (5). The terrain objects such tiles, bicycle pats, streets, parking lots, gardens, etc. situated in the same area are considered as well.

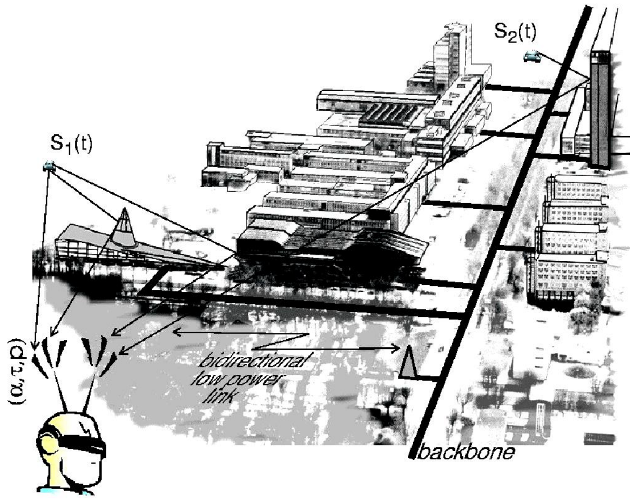


Figure 1: Schema of UbiCom augmented reality system

Figure 2: Test area: the campus of the Delft University of Technology

The selection of an appropriate approach is dependent on many factors such as required accuracy and detail, complexity of the reconstructed houses, availability and accuracy of data (images, maps), time and manpower constraints. Considering the relatively high accuracy requirements and the complexity of the buildings in the reconstructed area, we have chosen to apply manual and semi-automatic methods for reconstructing. Bearing in mind the normal behaviour of a walking person (i.e. looking mostly at objects at front and around), we concentrate on facades of the buildings and terrain objects rather than on the roof elements (usually not visible from street level). Thus, every building of interest is separately reconstructed as the main attention is on the front facades. In order to obtain the position of the individually reconstructed models in the real world and compute the precision of the reconstructed buildings, Least Squares Adjustment (LSA) is performed. Moreover, the reconstructed objects are to be organised in a 3D topological data structure (Zlatanova 2001). This requires clarifying of spatial relationships during the reconstruction procedure, which in some cases led to the development of supplementary software. The initial intentions were to limit the source data to only terrestrial (images taken from street level) and aerial images. At later stage, we have incorporated laser altimetry data and the large-scale digital topographic map (called GBKN). Thus the reconstruction procedure is a combination of several different approaches depending on the type of the objects and complexity of the shapes. Figure 3 gives a general view of the used software and data to obtain the 3D topologically structured model. We distinguish between procedures for reconstruction from terrestrial (Section 3) and aerial (Section 4) images, and procedures for reconstruction from laser data and GBKN (Section 5). The following sections discuss the major steps in more details.

3 RECONSTRUCTION FROM TERRESTRIAL IMAGES

In order to reconstruct the facades of the buildings, we follow four major steps:

1) collection of terrestrial images,
2) manual 3D reconstruction using the software package *PhotoModeller*,
3) accuracy assessment, and
4) recording in the 3D topological data structure.

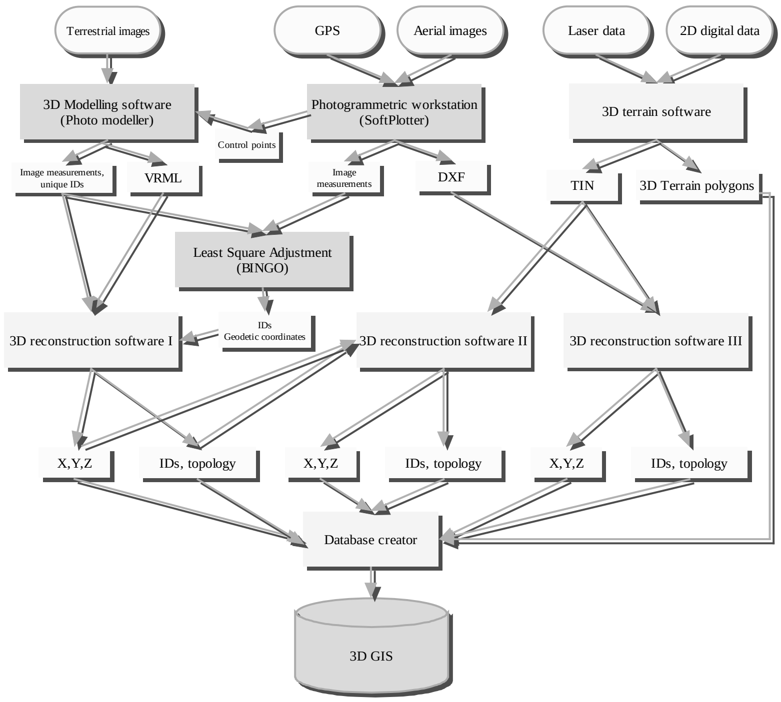
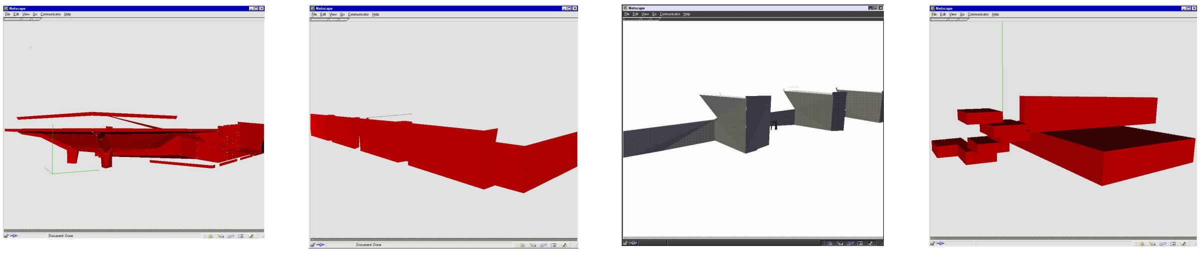


Figure 3: 3D reconstruction procedure

The images are taken with a handheld camera Kodak DCS420 (black and white) with resolution 1524x1012 and focal length 20 mm. In the process of reconstruction, more that 300 images are taken but actually less than 100 are considered appropriate for reconstructing (see Table 1). Most of the images are taken from street level and a small number by positioning the camera at the roof of the highest building in the campus. Since the angle between the images taken from this roof is relatively narrow, only a few of them are used.

The commercial modelling software package *PhotoModeller* is used for computing the 3D position of all measured points and reconstructing the surfaces. This approach involves a lot of manual work for selecting corresponding points on two and more images and connecting the points composing one triangle. It requires significant time and efforts to obtain the shape of a single building. However, since the measurements are performed and conducted by the operator, the reconstruction can be fully controlled. The output of the reconstruction is a 3D model in a model coordinate system. This model is then geo-referenced in *PhotoModeler* using control point co-ordinates derived from GPS and aerial photogrammetric measurements. The final co-ordinates of the 3D model in the national geodetic system are obtained by an integral LSA of all the measurements using software package *BINGO*. These are the photogrammetric measurements of the terrestrial and the aerial imagery, and control points measured with GPS. All the measurements on the aerial images are performed with the software package *SoftPlotter*.



a) b) c) c)

Figure 4: Images and reconstructed models of the buildings of: a) the Aula, b) the faculty of Applied Physics, c) the faculty of Mechanics and d) the Art

In-house software (see Figure 3: *3D reconstruction software I*) processes topologic data from *PhotoModeller* and geometric data from *BINGO* in order to obtain the correctly structured geo-referenced data. Unfortunately non-of the export file formats contains complete information about geometry and topology. Processing two export files from *PhotoModeller*, i.e. an ASCII file and VRML file, and one export file from *BINGO*, completes the structuring. The ASCII file contains the measured points with their IDs and 3D model co-ordinates. The VRML file contains the topology (description of the triangles) of all the surfaces reconstructed in *PhotoModeller*. According to the VRML syntax 1) the coordinates of the points (vertices of the faces in VRML) are stored only once in the description in VRML (in the *IndexetFacetSet* node), and 2) the faces can be grouped according to the objects they belong to. The *BINGO* file provides the geo-referenced coordinates of all the measured points. The in-house software unites coplanar triangles that share one edge in rectangular faces considering a given distance threshold. Thus, one or more rectangular faces represent every façade of a building (see Table 1).

Table 1: Project characteristics

3D object	Photos	Points (all)	Model points	Triangles	Rectangles
Aula	27	316	240	259	178
Physics	20	241	54	54	28
Mechanics	40	392	122	141	78
Art	9	64	56	98	50
All	96	1013	472	552	324

Table 2 shows the results of the LSA. The first three columns represent RMS of the object points in the local co-ordinate system. The good result, i.e. centimetre accuracy is indication for the appropriate selection of images and measured points. The second three columns show the co-ordinate precision resulting from the adjustment by *BINGO* in the national co-ordinate system. The observed decrease in the precision is basically due to lower resolution of the aerial images, and visibility and location of measured points used for geo-referencing. For example the Art monument is relatively small and white, surrounded by grass (i.e. dark object), which complicates the measurements due to the high contrast. The results given in bold represent the RMS precision of the objects after the common adjustment of all the measurements performed on all models. These results can be considered as an evaluation of the relative position of the individually reconstructed models. The improvement of the RMS precision achieved is due to rejection of some ambiguous measurements of control points.

Table 2: Average RMS precision values of the reconstructed buildings

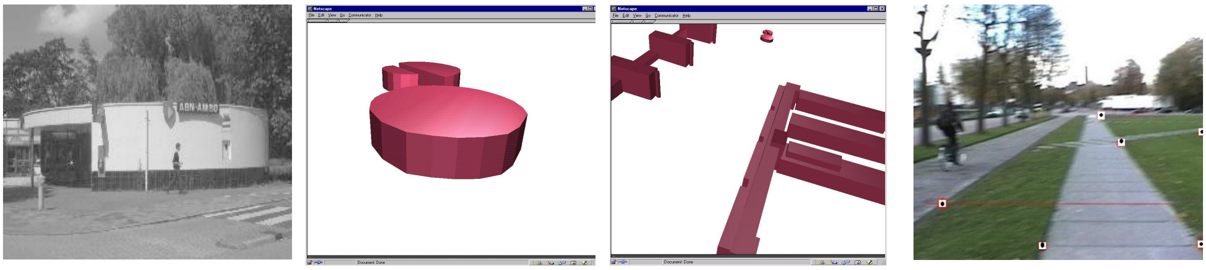
3D object	Model co-ordinates			Geo-referenced co-ordinates		
	RMSX (mm)	RMSY (mm)	RMSZ (mm)	RMSX (mm)	RMSY (mm)	RMSZ (mm)
Aula	23.3	8.9	19.8	77.7	76.5	76.7
Physics	34.8	14.3	36.3	71.8	67.6	73.7
Mechanics	20.7	11.1	25.8	152.0	98.1	84.7
Art	15.1	7.4	16.0	92.8	92.1	132.0
All				**67.1**	**66.1**	**77.5**

Clearly, using the procedures described above, we ensure the decimetre accuracy and the topological structuring required by the UbiCom augmented reality application.

Although applicable for reconstruction of buildings with complex shapes, this procedure might be rather slow and inefficient for particular types of objects. For example buildings with round shape, where no clear points can be referenced on multiple photos (see Figure 5a), or relatively simple long buildings where a lot of points are measured only to tie the images (see Figure 4 b, c). For such buildings we have adopted and extended a 3D semi-automatic procedure from aerial images originally developed at ITC, The Netherlands (Paintsil 1997).

4 RECONSTRUCTION FROM AERIAL IMAGES

The process of reconstruction from aerial images assumes that buildings have only vertical walls and non-over-hanging roof facets. This is to say, if roof outlines are projected vertically to the DTM, their footprints will be obtained. The procedure comprises a number of steps, i.e. DTM creation, manual digitising of roof outlines, processing of measurements for automatic reconstruction of walls and record in a database. The procedures for obtaining a DTM are described in the next section.

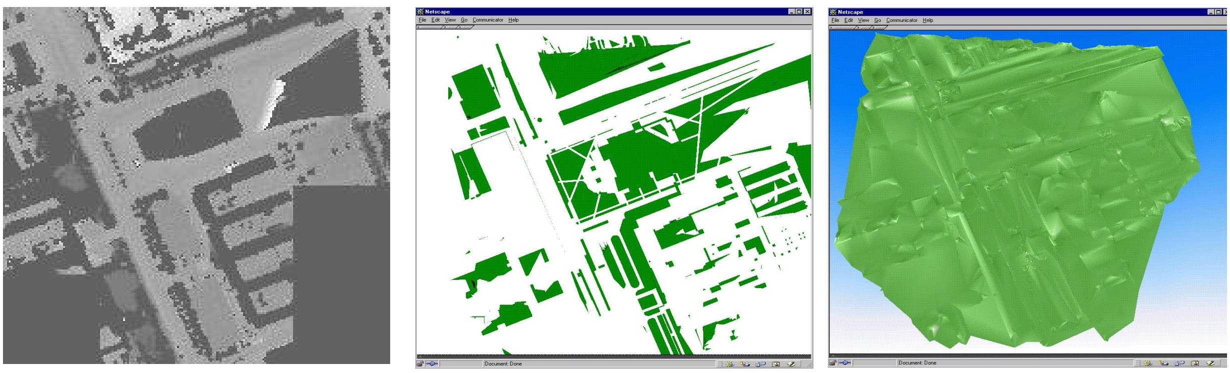


a) b) c) d)

Figure 5: The post office: a) image and b) model; c) gaps between two building parts; d) snapshot from a video camera

Once TIN is created, the process of reconstruction starts with the classical way of photogrammetric acquiring of data. Skeletal cloud of points is manually digitised following a number of restrictions depending on the type of the roof. Roofs of the same type are processed automatically as the complete 3D model of a building is created and automatically recorded in 3D topological data structure. The procedure is adapted for digital workstation *SoftPlotter,* i.e. new macros (i.e. scripts controlling the feature extraction) are developed in order to extract and export the measured points in a format appropriate for automatic reconstruction. The software called here *3D reconstruction software III* (see Figure 3) is responsible for:

- removing duplicated co-ordinates from the data extracted for each face,
- orientating faces in such a way that the corresponding normal vector points towards the outside of the building,
- generating building's footprints,
- constructing the roof and vertical walls of the building and
- creating 3D topology among the geometric primitives of the building.



a) b) c)

Figure 6: Reconstruction of terrain objects: a) filtered laser data set, b) GBKN, c) terrain objects in TIN

Figure 5b and Figure 5c show some of the reconstructed elements of buildings. The major advantage of the procedure (applicable for our goal) is the lack of gaps between the footprints and the terrain data. Once created, the footprints are used as constraints for re-triangulating TIN and incorporating them in the terrain. Furthermore, the 3D topology is created on the fly within the reconstruction process. The procedure is highly dependent on visibility of roof points on the aerial images. Invisible points or points on the same vertical line but on different levels disturb or make impossible the reconstruction. Figure 5c portrays an example of gaps between two buildings when the measured points on the roofs are at different level.

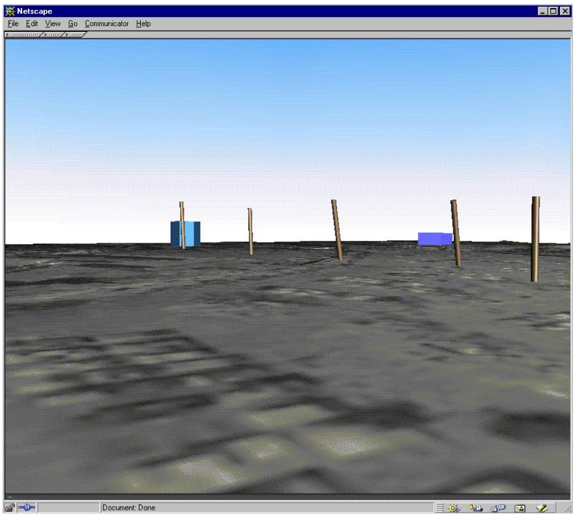
5 TERRAIN RECONSTRUCTION FROM LASER DATA AND THE GBKN

Different methods such as photogrammetry, surveying, etc., can be applied for the creation of a DTM. We experimented with 1) automatic DTM generation from aerial photographs, 2) manual digitising of terrain points and 3) combining 2D cadastral map with laser altimetry data. The automatic DTM generation is the one least suitable for densely built-up urban areas due to a number of reasons. The number of points is rather large, editing the automatically generated DTM is time-consuming, many non-terrain objects (i.e. cars and trees) are also included in the DTM and many terrain objects (e.g. streets, parking lots) cannot be distinguished as separate objects and have to be digitised anyway.

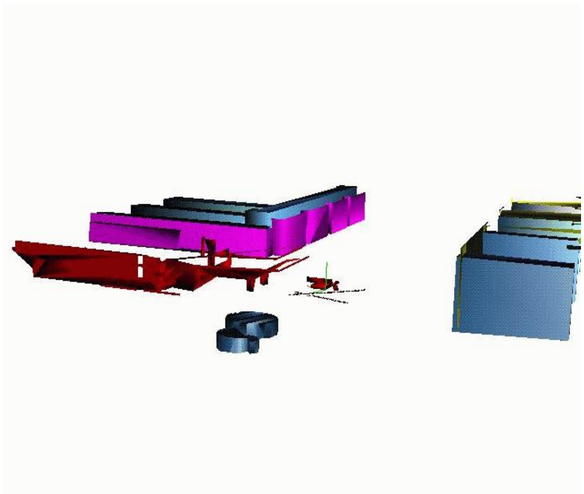
Creation of DTM from manually digitised points of terrain objects gives good results only if the density of objects is relatively high. Areas with large parking lots and gardens would require additional measurements of points inside the large objects.

The last approach requires minimum manual work and therefore it is used in our case. The available laser altimetry data have an average density of 5 points per square metre. They are further filtered to remove buildings, trees, cars and other non-terrain objects (see Figure 6a). More details about the filtering procedure can be found in Vosselman 2000. The laser points are used to create TIN. The TIN created in such way is quite similar in its structure to the automatically generated DTM, i.e. the points are distributed equally over the whole area where the filters are not applied. We "combine" the generated TIN with terrain objects from the topographic map of the Netherlands (i.e. GBKN) in order to incorporate the terrain objects (e.g. streets) into TIN. The terrain objects, which are stored as spaghetti lines in GBKN, are modified in such a way to be represented by polygons (see Figure 6b). For the purpose, the lines are extended until they intersect and the polygons are created using the intersection points (*ArcView*, *ESRI*). The 3D coordinates of the points are computed by interpolation of the z-coordinate on the TIN. Finally, a new TIN is generated using the terrain objects as constraints (see Figure 6c). More details about the TIN creation can be found in Gorte 1994.

While walking around, the mobile user might get into an area where only terrain objects such as parking lots, streets, bicycle paths, etc. are visible. Figure 5d shows an example of such an area with only a few buildings very distant from the user and practically not appropriate for positioning. In such cases, points of terrain objects, lamps and traffic signs are the only option. As mentioned above, we reconstruct terrain objects from GBKN and laser altimetry data (see Figure 3, 3D Terrain software). A modification of the procedure described in Section 4 allows reconstruction of lamppost and traffic signs (see Figure 7a).



a)



b)

Figure 7: Reconstructed models: a) lamps and b) all the reconstructed buildings

6 PROBLEMS IN THE 3D RECONSTRUCTION

At the end, the individually reconstructed objects are to be integrated in one topologically consistent 3D data model. The software referred to as *Database creator* (see Figure 3) cleans up duplicated points and controls the consistency of the faces composing one object.

We have encountered a number of problems in the reconstruction that required different solutions. The major problem refers to the measurements of points on the images. Invisible points make impossible the creation of faces and consecutively the correct reconstruction of the entire object. Some of the typical cases related to points in images at street level are listed bellow.

Invisible roof or corner points. Most of the roof points that have to be measured are badly visible or invisible due to higher distortion or some ornaments on the facades. This may be valid also for some corner points. New images from roofs of neighbouring buildings should bring an improvement. Another solution is the import of the incomplete model in CAD modelling software (Microstation, Maya, etc.) which requires development of software for correct transformation from the CAD export file formats to the 3D topological data structure. Intersection of already modelled planes under some assumptions such as planarity and orthogonality of facades can help in obtaining the missing points. An example of such complex building is the Aula (see Figure 4a). It has lots of corners and small overhangs that are difficult to get in two or three images with good intersection angles. Because of this the building has been a bit generalised. In particular this has been done on the staircases on either side of the building by closing some hollow spaces. The generalisation has been done in such a way that the real shape of the building was followed as much as possible.

Points measured in single images. Sometimes it is not possible to see points in more than one image. These points are measured in other images by using epipolar lines to reduce the finding of corresponding points to a 1D problem. The points measured in this way are not really independent measurements and the quality BINGO gives for these point will be too optimistic. This is especially the case around the staircases on either side of the building of the Aula.

Missing points on the ground level due to parked cars and other obstacles. In principle, very few points on ground level are accessible for measurements. Parked cars, bushes, etc. often occlude most of the points. This creates an integration problem when the buildings are combined with the terrain objects. The result is buildings “flying” over the terrain. Such cases we process automatically by in-house software called here 3D reconstruction software II. The points of the facades with the smallest z-coordinates are projected onto the ground to obtain the actual footprint. The software creates new faces using the original points and their projections, taking into account the correct orientation of the faces.

Difficulties to measure the same points on two and more images. Although the point seems visible on many images, it is sometimes impossible to measure exactly the same point. This effect is very strong for relatively small objects for modelling. For example the edges of the blocks of the Art monument are somewhat rounded. Since the object is rather small, the images are taken from close distance that makes the rounded edges very well visible. Therefore the operator experiences difficulty to find the same points in the different images.

Some of the objects can be modelled in PhotoModeller using predefined shapes, e.g. cylinders. Using these tools some parts of the Aula and the Departments of Mechanics are reconstructed. Unfortunately, such shapes cannot be processed in BINGO and they are excluded from the final model.

The reconstruction procedure based on aerial photo images revealed two significant problems, i.e. occluded points and multi-levelled points that can be projected onto the same point on the ground. Since new images are difficult to collect, the missing points have to be measured either with GPS or surveying methods. Gaps between different buildings can be removed only by manual editing of the points composing the faces.

Some of the objects for reconstruction have very long facades, which required utilisation of many images only for measuring tie points. This is especially true for the Department of Mechanics, and Physics (Table 1). Furthermore, since only the front facades are reconstructed from close-range photogrammetry, images only from one side of the buildings were used. All this caused large distortions in the shapes and the absolute orientation of the building models. Some model points were shifted with more than 2m. Control points at the two ends of the building improve the results significantly.

Yet the reconstruction of complex roofs has to be resolved. Besides the trivial manual measuring and reconstruction, an attractive solution may be the utilisation of laser altimetry data. The points of the roof facets belonging to one plane (within certain threshold) can be filtered and the equation of the plane can be computed. Intersecting the planes, the faces of the roof can be obtained. More information on such procedures can be found in Maas and Vosselman 1999 for reconstructing from laser data and in Schmidt 1989 for reconstructing from images.

The integration of 3D models created from different data sources is still waiting for a fully automatic solution. For example, combining roof faces reconstructed from aerial images and walls reconstructed from terrestrial images requires manual interaction. The most critical points are at the roofs and the footprints of the buildings. The problem is twofold, i.e. gaps can be obtained both due to measurement errors and due to invisibility of points. In the first case, a procedure for finding the most probable position of the doubtful point has to be designed. In the second case, new faces have to be constructed.

7 CONCLUSIONS

We presented our work on 3D reconstruction of an outdoor environment for augmented reality applications. We applied different approaches for reconstruction in order to ensure the required accuracy and organisation of data. The work on the reconstruction of the buildings in the campus proved once again that one optimal procedure for reconstruction does not exist. Complexity of real objects varies in such large extends that manual, automatic and semi-automatic methods may need to be utilised in order to achieve accuracy requirements, reduce the manual work and organise the data according to the specifications. Furthermore, limitation to only few data sources may require extra efforts to collect data and perform measurements. Therefore, we recommend careful survey and analysis of all the possible data sources and their processing before the reconstruction is started. However, one should have in mind that many data sources and the combination of different approaches always creates integration problems.

We consider our work a successful attempt to measure, reconstruct and structure real objects in a manner suitable for mobile augmented reality applications:

1) the reconstruction presented supplies appropriate 3D data with sufficient accuracy for rendering of virtual objects,
2) the 3D reconstruction software integrates triangles into more intuitive shapes for facades, i.e. rectangles, and thus facilitates the storage of 3D line features,
3) the reconstruction software provides automatic solutions to some of the reconstruction problems discussed above,
4) the reconstruction procedures are linked by a number of programs that allow 3D topology to be automatically built.

References

Gorte, B., 1994, Tools for advanced image processing and GIS using ILWIS, ITC publication 24, Enschede, The Netherlands

Maas, H.-G. and G. Vosselman, 1999, Toepassingmogelijkheden van Vliegtuig-Laseraltimetrie, in: Geodesie zonder grensen, Delft University Press, ISBN 90- 407- 1950- 0, pp. 23-29.

Paintsil, J.,1997, 3D Topographic Data By Aerial Photogrammetry, MSc thesis, ITC, Enschede

Pasman, W. and F.W. Jansen, 2001, Realistic low-latency mobile AR rendering, Proceedings of VAA01, 21-22 June, Trinity College, Dublin, Ireland, pp. 81-92

Persa, S and P. Jonker, 1999, On Positioning Systems for Augmented Reality Applications, Handheld and Ubiquitous Computing 1999, Springer Lecture Notes in Computer Science 1707

Ubicom project, 2000, available on http://bscw.ubicom.tudelft.nl/

Schmidt, W.F., 1989, Segmentation and analysis of range images, MSc thesis, the Delft University of Technology, The Netherlands, 54 p.

Vosselman, G, 2000, Slope based filtering of laser altimetry data, International Archives of Photogrammetry and Remote Sensing, XXXIII, B3/2, pp. 935-942

Zlatanova, S., 2001, 3D modelling for augmented reality, Proceedings of ISPRS, Vol.34, Part 2W2, "Dynamic and multi-dimensional GIS", 23-25 May, Bangkok, Thailand, pp. 215-420