Mobile mapping and computer vision for generation of 3D site models

F. Leberl
Institute of Computer Graphics and Vision,Graz University of Technology, Austria

K.F. Karner
VRVis Research Center, Graz, Austria

ABSTRACT: Cultural heritage documentation has had spectacular successes in the preservation and restoration of sites destroyed by natural disasters or by war. These successes were largely based on photographic documentation, that were translated into construction information using manual methods to include all types of 3-dimensional cues, be this stereo, shadows, shading or the simple use of geometric knowledge about a building and combining this with a single image. Computer vision and direct shape measurements now add the option for an automated analysis of the imagery, thereby enabling a more complete assessment of sites than ever before. We review a series of projects and efforts to develop technologies, systems and applications with site models at accuracies in the range of at least ± 10 cm.

1 INTRODUCTION

The intersection of computer vision, computer graphics and photogrammetry is very evident in the application to city models, in processing and using models of buildings, of the insides of buildings and of their objects. This has been intensively studied since about a decennium. Early work was largely manual (Gruber et al., 1995) or machine supported (Lang & Schickler, 1993), and inspired the use of video streams and uncalibrated images (Faugeras, 2001, Hartley & Zisserman, 2000). The idea of the "Cyber City" was developed, and this has even left the realm of academia and has led to commercial activities, with the major markets being the heritage documentation and telecommunication industry.

A series of European Union projects focused on automating the creation of models of archeological and historical sites and led to the use of a variety of shape-modeling approaches, be this from video streams, from blocks of calibrated as well as uncalibrated images, from single images, or from direct scanning.

We report in this contribution on a steady stream of activities that we have come to denote as "Virtual Habitat". The project context was mostly the interest in historical sites, but also city models for telecommunications and the general purpose of an urban 3-D geographic information system for engineering, planning, citizen participation, disaster preparedness, armchair tourism etc. "Heritage mapping" is not a separate field but part of the mapping of urban habitats.

2 AERIAL SITE MODELING

Currently, aerial photogrammetry is undergoing a "paradigm shift" (Leberl & Thurgood 2004) which means the transition from minimizing the number of film photos due to human operator intensive processing to maximizing the robustness of automation due to high redundant image information using new large format digital aerial cameras.

In our aerial workflow we use images from an UltraCam-D camera from Vexcel Imaging which delivers 16 bit pan-sharpend RGB-NIR images with a size of 11500 x 7500 pixels. The camera is able to deliver images at intervals down to 1 sec.

Our workflow includes the following steps: a classification of all images, the aerial triangulation (AT) using area and feature based POIs, a dense matching to generate a dense DSM (digital surface model), a refined classification using the DSM, a ‘true’ orthophoto production, and the estimation of a DEM. In this part of the paper we will focus on the automatic AT and on dense matching only.

2.1 *Automatic Aerial Triangulation*

Digital airborne cameras are able to deliver high redundant images which result in small baselines. Normally, the strips of images should have at least 80% forward overlap and a minimum of 20% side overlap (in urban areas 60% side overlap). This high redundancy - one point on the ground can be seen in up to 15 images - and the constraint motion of a plane help to find good starting solutions needed for a fully automated AT. Nevertheless, an accurate extraction of tie points is needed for a robust and accurate AT (Thurgood et al. 2004). Our POI extraction is based on Harris points and POIs from line intersections (Bauer et al. 2004).

The POIs from line intersections which we call ‘zwickels’ are very suitable for urban areas. Zwickels are sections defined by two intersecting line segments, dividing the neighborhood around the intersection point into two sectors.

After the POIs extraction in each image we calculate feature vectors in the close neighborhood. Feature vectors are used to find 1 to n correspondences between POIs in two images. Using affine invariant area based matching the number of candidates is further reduced. For all remaining candidates we iteratively apply an affine transformation to maximize the cross-correlation score. As a result we get a list of corresponding points. In order to fulfill the non-ambiguous criteria, only matches with a high distinctive score are retained. The robustness of the matching process is enhanced by processing a back-matching as well.

Another restriction is enforced by the epipolar geometry. Therefore the RANSAC method is applied to the well known five point algorithm (Nister 2003). As a result we obtain inlier correspondences as well as the essential matrix. By decomposition of the essential matrix the relative orientation of the current image pair can be calculated.

This step is accomplished for all consecutive image pairs. In order to get the orientation of the whole set, the scale factor for additional image pairs has to be determined. This is done using corresponding POIs available in at least three images. A block bundle adjustment refines the relative orientation of the whole set and integrates other data like GPS, DGPS, IMU or ground control information. Figure 2 shows an oriented block of 7 X 50 aerial images together with the used 3D tie points on the ground. The whole block of images was processed without any human interaction.

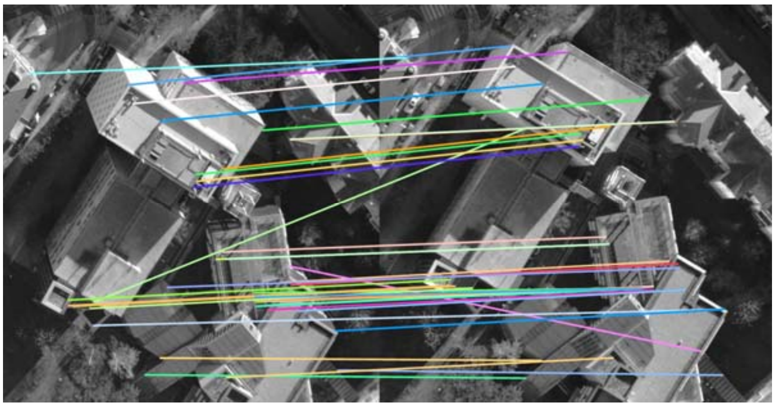


Figure 1. In this example we use zwickels only to show their strength in urban areas. There are only two outliers within the best 25 matches before the epipolar constraint is applied. Corresponding POIs are connected by lines.

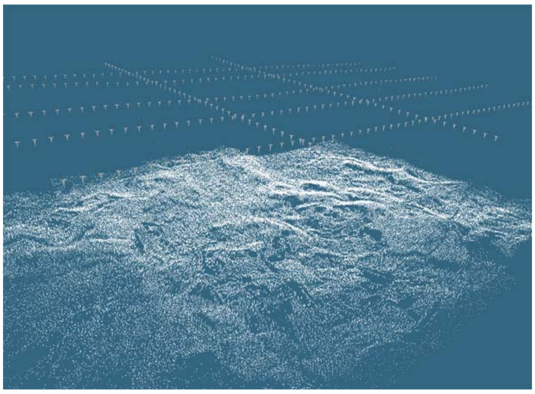
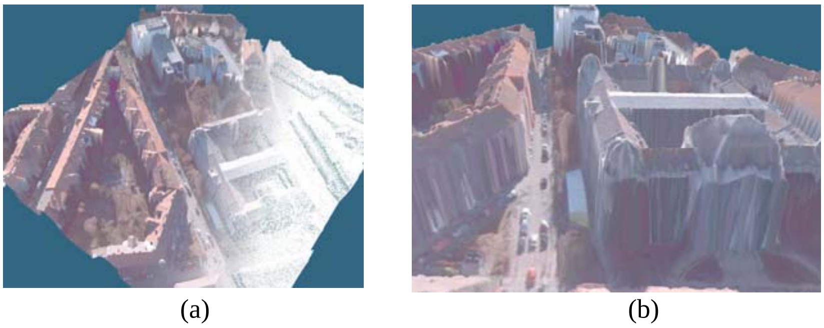


Figure 2. 7 strips of about 50 images each (5 strips flown east-west and 2 north-south) denoted by small arrows are oriented to each other using about 70.000 tie points on the ground which are shown as white dots.

2.2 *Dense Matching*

Once the AT is finished we perform a dense area based matching to produce a dense DSM (digital surface model). Recent years saw more new dense matching algorithms were introduced. A good comparison of stereo matching algorithms is given in a paper by Scharstein et al.(2002). Recently, a PDE based multi-view matching method was introduced by Strecha et al. (2003). In our approach we focus on an iterative and hierarchical method based on homographies to find dense corresponding points.

Figure 3 shows some results of our approach.



(a) (b)

Figure 3. A dense triangular mesh is calculated from oriented aerial images. a) Only parts of the mesh are textured to see the high geometric resolution of the mesh. b) A second view of the same mesh from close above the roofs. Due to a flight height of about 1000m above ground, the facades cannot be modelled very well due to small intersection angles. The "drawing" of the texture is from a 2D ortho image and therefore is not addressing the focades.

2.3 *Orthophoto*

A 'true' orthophoto is obtained by orthoprojection of the DSM (see Figure 4). The color information of the orthophoto is calculated using all available aerial images and is based on view-dependent texture mapping described in (Bornik et al. 2001).

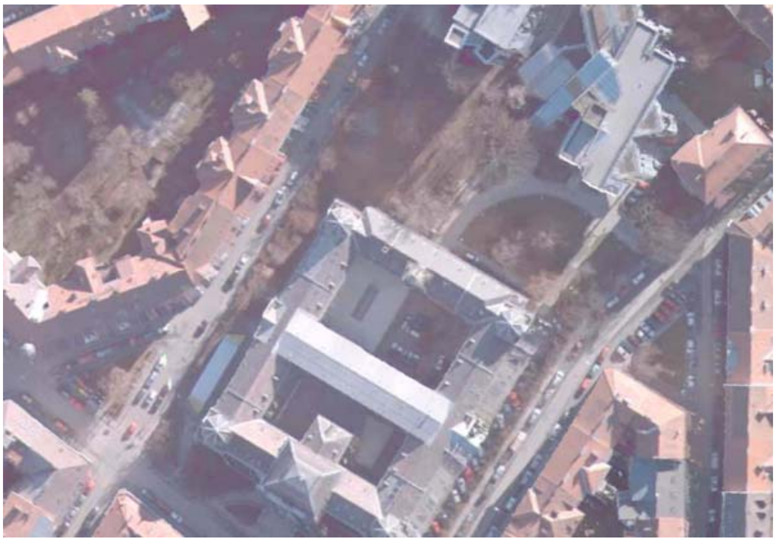


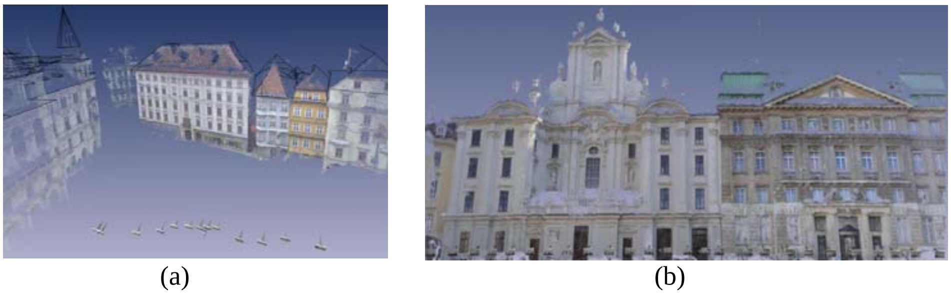
Figure 4. 3D view of the textured DSM with the projection mode set to orthonormal projection and a vertical viewing direction. The facade surfaces are not visible, as expected from a ‘true’ orthophoto.

3 TERRESTRIAL SITE MODELING

A commercial context led to the development of an entire “system” called MetropoGIS to go from the collection of terrestrial images and laser scanning point clouds to interpreted models of urban landscapes. In fact, the idea of the “Virtual Habitat” is at the core of a commercial development to create 3D input or “content” for an urban 3D GIS. A robust approach for the orientation of image sequences showing mainly man-made structures are explained in (Klaus et al. 2002). In this Section we concentrate on the modeling step only.

3.1 *Dense matched surface models*

The procedure applied here is very similar to the one explained in Section 2.2. Two results using this approach are shown in Figure 5.



(a) (b)

Figure 5. a) Two compounded geo-referenced sequences of the Grazer Hauptplatz augmented with roof lines. b) Fully automatic result from a sequence of 26 images acquired at ´Am Hof´ in Vienna.

3.2 *Feature-Based Modeling*

The set of line segments per image together with the known orientation of the image sequence are the input for the line matching algorithm. Figure 6 shows the extracted 2D lines in one of the input images. Our approach closely follows the one described by Schmid and Zisserman (Schmid & Zisserman 2000). The result of the line matching process is a set of 3D lines in object space. Basically the algorithm works as follows: For a reference line segment in one image of the sequence potential line matches in the other images are found by taking all lines that are enclosed by the epipolar lines induced by the endpoints of the reference line segment. Each of these potentially corresponding line pairs gives a 3D line segment (except for those, which are parallel to the epipolar line, since in this case no intersection between the epipolar line and the image line can be computed). The potential 3D lines are then projected into all remaining images. If image lines are found which are close to the reprojection, the candidate is confirmed, else it is discarded. Finally a correlation based similarity criterion is applied to select the correct corresponding line. Figure 7 shows two views of the extracted 3D line set. Obviously, due to the small vertical baseline the geometric accuracy of the horizontal line segments is limited. A more detailed description of the involved steps can be found in (Bauer et al. 2002).

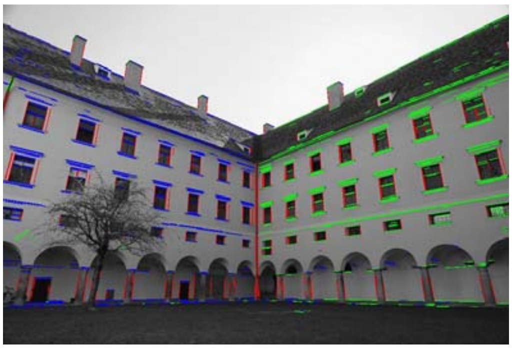


Figure 6. Extracted lines in of the input images. Automatically extracted vanishing points are used to improve and to classify the 2D line segments.

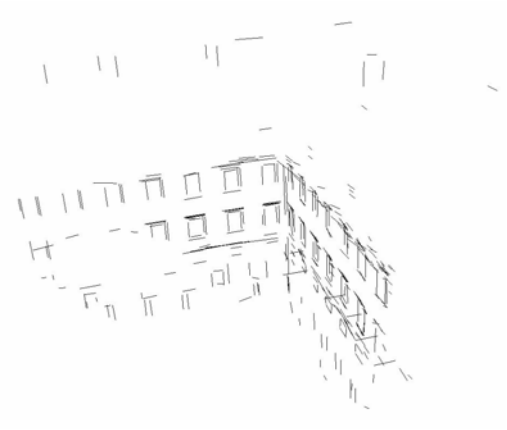


Figure 7. A views of the 3D line matching results.

3.3 *Using Curves and Planes to Analyze Point Clouds*

Schindler & Bauer (2003) developed an approach for the replacement of point clouds by curves and planes. Figure 8 illustrates the basic idea by means of a historical structure. The points a plane can be assembled from the stereo parallaxes. The application of knowledge can map entire facades, windows, arches.

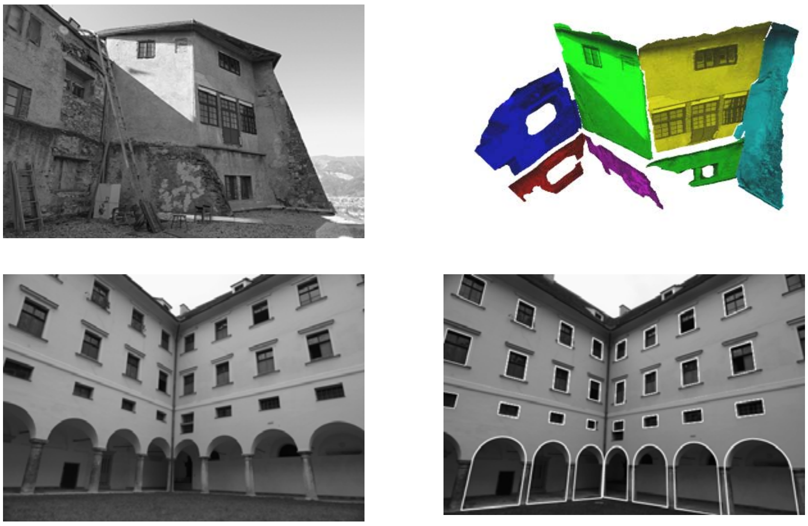


Figure 8. Input image representing a 3D point cloud (upper left) and individual textured planes of the structure (upper right). Input image for curve reconstruction (lower left) and curves extracted automatically (lower right).

4 GPU-BASED MATCHING

Since dense matching using high redundant image information is computationally expensive, we have investigated several possibilities to speed it up using modern graphics hardware. Our approach generates a 3D mesh for objects visible in a pair of stereo images by finding a dense set of corresponding points between the images (see (Zach et al. 2003) for a more detailed description of our approach).

The basic element of our reconstruction software is image warping.

- Most stages of dense stereo matching can be effectively executed in parallel. Performing these steps with a general purpose CPU does not exploit the parallelism.
- In addition, for some elementary operations required during the stereo matching procedure (like bilinear pixel access) modern 3D graphics hardware has very fast special purpose circuits.

For these reasons we moved as much of the stereo matching procedure as possible to modern graphics hardware. With the general availability of programmable vertex and pixel shaders we are able to execute almost all stages of our stereo matching method on 3D hardware without the need of transferring a large amount of data between main and video memory.

We achieve a stereo model from 1K X 1K imaging in about 1 sec.

5 TEXTURING

Once the 3D modeling is created we automatically generate a texture which is most consistent with the available input images. Attention is paid to the actual generation of the texture where the three dimensional model and 2D-parameterization is already known and only the surface color properties are missing (our implementation is based on (Lvy et al. 2002).

This is achieved by first segmenting the model, then mapping the three dimensional geometry to the planar texture image which is afterwards filled by projecting the corresponding point in the three dimensional domain into all available images. The resulting values are used to compute a high quality, robust texture considering the visibility, view angle and footprint of the projection.

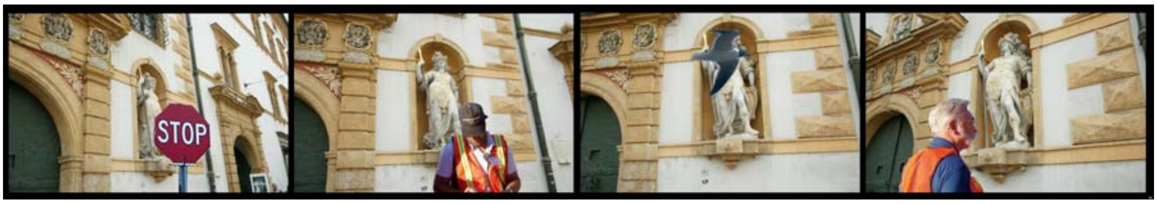


Figure 9. Selection of four input images from a set of seventeen taken in the center of Graz with a calibrated camera. The focal length, aperture as well as the exposure time were kept constant.

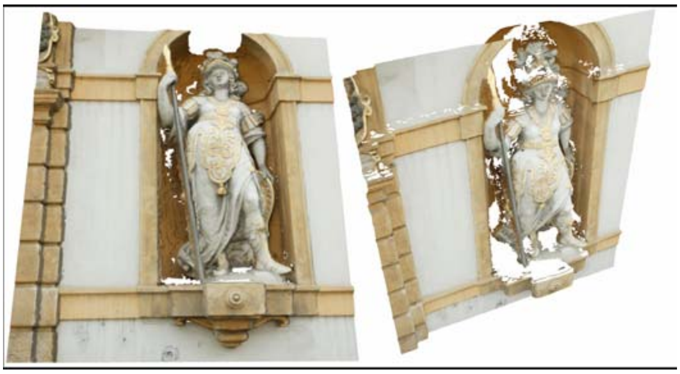


Figure 10. The model with a texture synthesized from all available
seventeen input images. Regions which are not visible in any view
can not be reconstructed and therefore also have no texture.
Examples of such holes are shown in the image on the right hand side.

Note that certain images show occlusions. These get eliminated by means of the redundancy from multiple images. The various studies and developments of tools to model buildings from images has recently led to an involvement in European initiatives.

6 MODELING EUROPEAN HISTORICAL BUILDINGS: THE CULTURE2000 PROJECT

Is a project to use novel interactive modeling systems to create a detailed and complete 3D reconstruction of cultural heritage from a set of photographs. Our approach utilizes terrestrial images of facades, taken by a hand-held digital consumer camera using short baselines. Aerial images are used to model parts of the scene (e.g. roofs) which are not visible from the terrestrial images, and they provide the geometric reference. The relative orientation of the terrestrial photographs is calculated automatically, whereas the integration of the aerial images is performed with minimized human interaction using methods described above. Once we have determined the orientation of all images, we are able to extract 3D information from the image sequence automatically, by employing different area and feature based matching techniques resulting in 3D points, lines and surfaces as described. A human operator to use his unique interpretation and segmentation abilities. The modeling system uses a 2D segmentation and interpretation whereas the system is responsible for 3D modeling. The user interface as monocular and 3D (see Figure 11). As a result of the modeling process we are able to obtain a coarse as well as a highly detailed 3D model of the scene. The goal of the creation exploring cultural heritage we visualize the created virtual model in a virtual reality environment (see Figure 12), much along the lines of tourism.

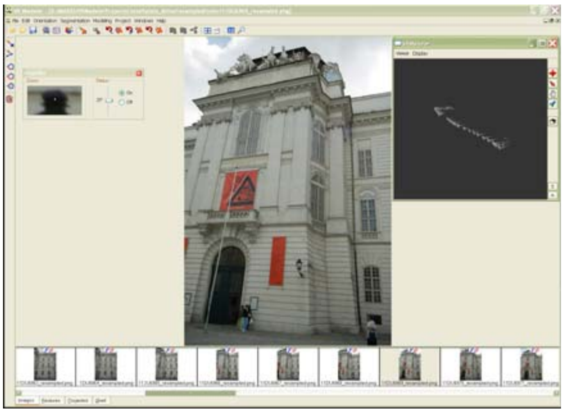


Figure 11. User interface, which includes a magnifier, an image viewer, a 3D viewer and an image preview box, all for monocular interaction with a multi-image data set.

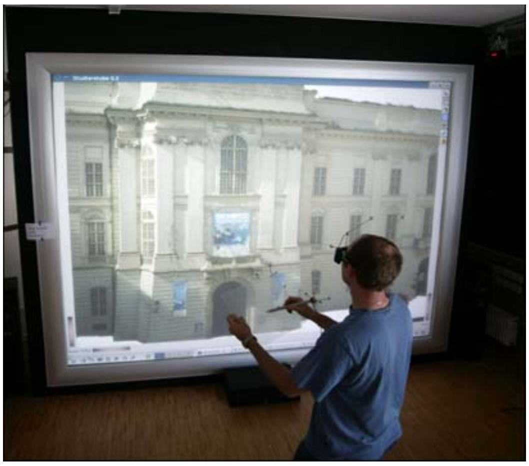


Figure 12. Interaction with a 3D model obtained from Culture 2000 and its monocular interactive approach of Fig. 11.

7 DATA BASE AND VISUALIZATION: THE STRAPAMO PROJECT

A high quality visualization system for changing urban environments is a powerful tool for city planners. In order to accommodate the time-dependent nature of a city in such a system, we have built a four dimensional information space which facilitates both realtime rendering and database management. Three dimensions describe space, the forth dimension addresses time. Our system consists of a Database Server, which solves GIS-related problems, and a fast visualization tool called AveViewer based on the Advance Visualization Engine (AVE) developed at the VRVis Research Center.

7.1 *Database System*

The basis for data exchange between the viewer and the database, as well as for importing data, is VRML – an open 3D graphics format. The database server, utilizing MySQL 4.1 satisfies the OpenGIS standards, and facilitates the management of large amounts of structured geometric, raster and metadata content extended by timestamps. AveViewer uses this time-dependent information for a number of unique visualization methods, which are especially suited for urban planning.

The Database Server stores the content intended for fast rendering in the arrangement most appropriate for instant reaction to the AveViewer demands. On the other hand, it organizes data transparently in order to provide multipurpose interface for common GIS analysis. Core of the Server thus is a well-arranged spatial database accessible by geometrical, quantitative, metadata and other SQL queries (most common query language). The main dataflow diagram of the solution is shown in Figure 13.

7.2 *Database AVE Viewer*

The final, and most visible part of the system is the viewer to navigate cities in both space and time. It acts as a client to the Database Server and requests geometry and metadata on demand. In order to reduce communication overhead it also incorporates a cache that is being filled with the results of all geometry requests. Thus the network time delay only appears the first time when a specific building or object in the database is requested for viewing.

The viewer has implemented with the Advanced Visualization Engine and provides a realtime rendering functionality for walk-throughs and fly-overs of complete cities modeled at the medium-detail level including the typical few instances of fine-detail models. A screenshot captured during a flight over Graz is shown in Figure 14.

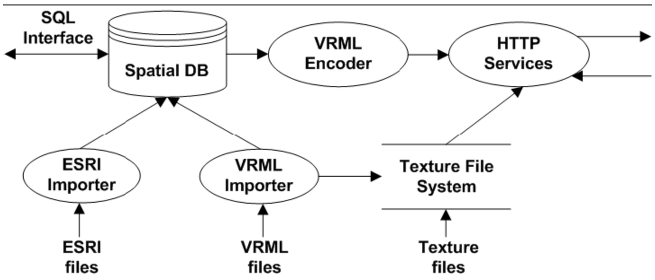


Figure 13. Main context of the MetropoVis Database Server dataflow for the Strapamo project.

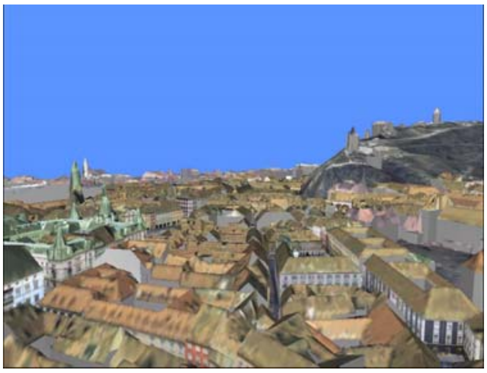


Figure 14. Screenshot of a a Fly-over of the city of Graz.

It serves as a test-bed for implementing different functionality based on the desired application area. New functionality of the viewer has been built on the basis of a visualization kernel with features that are important for city modeling, such as secure transmission (Hesina & Tobler 2003) and optimized texture management (Hesina et al. 2004).

8 CONCLUSION AND PERSPECTIVE

The 3D modeling of urban spaces has increasingly become a “killer application” for computer vision research. Data sources were traditionally metric aerial and ground-based images. More recently, this has been extended to include non-calibrated images from consumer cameras, non-calibrated video-streams, laser scanners. The 3D data product initially is a point cloud that gets then subjected to an automated interpretation.

We report in the contribution on a range of projects to increase the bowl of automation in creating 3D models of the urban environment. Our intention are buildings, facades, statues and the evolution of objects over time. Our report addresses work that has come about in a fruitful co-operation of two allocated entities: A team of the VRVIS Research Company and an Institute of Graz, University of Technologies.

ACKNOWLEDGEMENTS

We would like to thank Joachim Bauer, Barbara Gruber, Gerd Hesina, Andreas Klaus, Mario Sormann, Christopher Zach and Lukas Zebedin for their contributions to this paper. The presented work has been done in the VRVis research center, Graz and Vienna/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus.

REFERENCES

Bauer J., Bischof H., Klaus A. & Karner K. 2004. *Robust and fully automated Image Registration using Invariant Features*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 35, Part B3 :1682-1777.

Bauer J., Klaus A., Karner K., Zach C. & Schindler K. 2002. MetropoGIS: *A Feature based City Modeling System*. Photogrammetric Computer Vision 2002 (PCV'02).Vol. XXXIV part 3B, pp22-27, ISPRS - Commission III, Symposium 2002. September 9 - 13, 2002. Graz, Austria.

Bornik A., Karner K., Bauer J., Leberl F. & Mayer H. 2001. *High-quality texture reconstruction from multiple views*. The Journal of Visualization and Computer Animation 12(5): 263-276. John Wiley & Sons, Ltd.

Faugeras O. 2001. MIT Press. 646 p.

Gruber M. & Meissl S. & Böhm R. 1995. *Das dreidimensionale digitale Stadtmodell Wien*. Erfahrungen aus einer Vorstudie. VGI (Österreichische Zeitschrift für Vermessung und Geoinformation), Vol. 1+2, pp. 29-36.

Hartley R.& Zisserman A., 2000. *Multiple View Geometry*. Cambridge University Press, 642 p.

Hesina A. & Tobler G. R. F. 2003. *Secure and Fast Urban Visualization* Proceedings of the CORP - GeoMultimedia 2003, WebDownload, pp. 231-234, Vienna. Austria.

Hesina A.& Maierhofer G. & Tobler R. F. 2004. *Texture Management for high-quality City Walkthroughs*. Proceedings of the CORP – GeoMultimedia 2004 pp. 305-308. Vienna. Austria.

Klaus A.& Bauer J. & Karner K.& Schindler K. 2002. MetropoGIS: *A Semi-Automatic City Documentation System*. Photogrammetric Computer Vision 2002 (PCV'02). Vol XXXIV part 3A, pp. 127-192, ISPRS - Commission III, Symposium 2002. September 9 - 13, 2002. Graz, Austria.

Leberl F. & Thurgood J.. 2004. *The Promise of Softcopy Photogrammetry Revisited*. ISPRS 2004. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XXXV ISSN 1682-1777. Istanbul, Turkey.

Lang F. & Schickler W. 1993. *Semi-automatische 3D-Gebäudeerfassung aus digitalen Bildern*. Zeitschrift für Photogrammetrie und Fernerkundung, Vol.5, pp. 193-200.

Lvy B.& Petitjean S. & Ray N. & Maillot J.. 2002. *Least squares conformal maps for automatic texture atlas generation*. In SIGGRAPH ’02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques, pages 362–371. ACM Press.

Nister D. 2003. *An efficient solution to the five-point relative pose problem*. CVPR 2003, pages II: 195–202.

Rigaux P.& Scholl M. O. & Voisard A. 2001. *Spatial Databases: With Application to GIS*, Morgan Kaufmann, 1st edition.

Scharstein D. & Szeliski R. 2002. *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. IJCV, 47(1/2/3):7-42.

Schindler K. & Bauer J. 2003. *A model-based method for building reconstruction*. Proccedings ICCV Workshop on Higher Knowledge in 3D Modelling and Motion Analysis, pp. 74-82. Nice. France

Schmid C. & Zisserman A. 2000. *The geometry and matching of lines and curves over multiple views*. IJCV, 40(3):199–233

Sormann M. & Klaus A.& Bauer J. & Karner K. 2004. *VR Modeler: From Image Sequences to 3D* Models. SCCG (Spring Conference on Computer Graphics) 2004. ISBN 80-223-1918-X, pg. 152-160.

Strecha C.& Tuytelaars T.& Gool L. Van. 2003. *Dense Matching of Multiple Wide-baseline Views*. ICCV 2003, vol 2, pp. 1194-120.

Thurgood J.& Gruber M. & Karner K. *Multi-Ray Matching for Automated 3D Object Modeling*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume XXXV. ISSN 1682-1777. Istanbul, Turkey.

Zach C. & Klaus A. & Hadwiger M. & and K. Karner. 2003. *Accurate dense stereo reconstruction using graphics hardware*. In EUROGRAPHICS 2003, Short Presentations, pp. 227-234.

MySQL 4.1 WebDownloads, http://dev.mysql.com/downloads/mysql/4.1.html

OpenGIS® Simple Features Specification for SQL, WebDownload http://www.opengeospatial.org/docs/99-049.pdf

MetropoVis project home page, WebDownload http://www.vrvis.at/research/projects/MetropoVis/