FAST AND DETAILED 3D RECONSTRUCTION OF CULTURAL HERITAGE

Mario Sormann[a], Bernhard Reitinger[b], Joachim Bauer[a], Andreas Klaus[a], Konrad Karner[a]
[a] VRVis Research Center, Inffeldgasse 16/II, A–8010 Graz, Austria
[b] Institute for Computer Graphics and Vision, Graz University of Technology,
Inffeldgasse 16/II, A–8010 Graz, Austria
sormann@vrvis.at

KEY WORDS: orientation, 3D reconstruction, segmentation and interpretation, virtual reality

ABSTRACT

The inner cities of Graz and Vienna were awarded World Cultural Heritage by UNESCO and in the year 2003 Graz was the Cultural Capital of Europe. Therefore we report in this paper on a novel interactive modeling system to create a detailed and complete 3D reconstruction of city centres from a set of photographs. We emphasize in our approach on utilizing terrestrial images of facades, taken by a hand-held digital consumer camera using short baselines. Aerial images were used to model parts of the scene (e.g. roofs) which are not visible from the terrestrial images. The relative orientation of the terrestrial photographs is calculated automatically, whereas the integration of the aerial images is performed with minimized human interaction. 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. Due to the fact that we concentrate on a complete 3D reconstruction of city centres we apply several fusion techniques to the different extracted model representations. In particular, we decided to support this task by a human operator in terms of his unique interpretation and segmentation abilities. The modeling system is efficient and easy to use because the user can focus on the 2D segmentation and interpretation whereas our system is responsible for the 3D modeling. Therefore we developed the user interface as a monocular 3D modeling system. As a result of the modeling process we are able to obtain a coarse as well as a high detailed 3D model of the scene. Taking into account the goals of rehabilitation of city centres we visualize the created virtual model in a virtual reality environment.

1 INTRODUCTION

The rehabilitation and preservation of cities, especially city centres with historic sites is an important problem and often an expensive and tedious process even with modern photogrammetric and computer vision techniques. In general city centres are often subject of erosion since they have gone through many phases of construction, damage and repair. Therefore it is important to have a detailed and accurate reconstruction of such sites. Furthermore it is an attractive application to visualize and explore cultural heritage architecture in a virtual reality environment.

Our research aims to integrate these requirements into our 3D modeling system, which allows a user to reconstruct a geometric model of a historic site from a set of photographs more accurate and convenient than current available computer vision approaches. The scene is captured directly on site with a hand-held digital consumer camera using short baselines. After some preprocessing and acquiring the relative orientation as described by Nister (Nister, 2003) and Horn (Horn, 1990) between all image pairs we reconstruct a coarse as well as a high detailed geometric model of the historic site with our intuitive interactive 3D modeling system. Basically, we combine automatic area and feature based modeling techniques with a minimum on human interaction to obtain a consistent geometric model of the scene. Additionally we demonstrate the accuracy and quality of the obtained reconstructed model in an virtual reality environment.

The proposed system has been developed and successfully tested within the *Creative Histories - The Josefsplatz Experience* project. An overview of the Josefsplatz is presented in Figure 1. The goal of the Creative Histories project is the as-built documentation and reconstruction of a complex part of a city centre (Josefsplatz), the 3D modeling from historical pictures and paintings and the modeling of historical textures and events. In this paper we will concentrate on the as-built documentation and reconstruction.

Figure 1: Overview of the Josefsplatz in Vienna.

The rest of the paper is organized as follows: Section 2 provides an overview of related work. Our interactive modeling system for fast and detailed reconstruction of city centres is presented in Section 3. The visualization and exploration of the city centre in a virtual environment is sketched

in Section 4. Some results to illustrate our modeling approach are presented in Section 5. Section 6 concludes our approach and gives some ideas of future work.

2 RELATED WORK

The reconstruction of 3D models from 2D images has a long tradition within the photogrammetric community. There are two classes of research fields related to our method: automatic reconstruction and human assisted reconstruction from image information.

2.1 Automatic Reconstruction

The automatic reconstruction of 3D models can be separated into two different modeling approaches: area based modeling and feature based modeling.

Area based modeling methods are based on an estimation of dense 3D point clouds from image sequences (Pollefeys et al., 2000). The critical part behind these methods is to robustly find corresponding points within the image sequences. From the corresponding points the relative orientation of each image can be estimated (Hartley and Zisserman, 2000). This procedure is applied to many pixels within stereo or multiview images which results in dense 3D point clouds. Frequently a continuity constraint is used to allow smooth reconstructions. A good collection and comparison of different stereo matching methods can be found in (Scharstein and Szeliski, 2002). A recently proposed area based modeling approach (Strecha et al., 2003) utilizes partial differential equations (PDE's) for dense depth extraction from multiple images.

Feature based modeling approaches are widely used in modern computer vision systems. They reduce the amount of data to be processed, and also increase the robustness and accuracy of measurements in digital images. A general overview of feature based modeling methods is given in Baillard et al. (Baillard et al., 1999), where they propose a line matching method over multiple oriented views.

Our research aims to combine feature and area based modeling techniques to obtain a complete and a consistent 3D model from an image sequence.

2.2 Human Assisted Reconstruction

Many of the current 3D reconstruction systems are based on human input at some point of the reconstruction process. A very well known concept in this research field is the modeling and rendering architecture proposed by Debevec (Debevec, 1996). This 3D reconstruction system, called *Façade* is separated into two main modeling steps. The first step allows the recovery of a basic geometry model, whereas the second one represents a view dependent texturing system to improve the geometric details of the basic model. The popular commercial product *Image Modeler* from RealViz (ImageModeler, 2004) is inspired by the above described system. Another well known method proposed by Shum (Shum et al., 1998) presents an interactive modeling system that constructs 3D models from a collection of panoramic image mosaics.

3 3D RECONSTRUCTION OF CULTURAL HERITAGE

In this section we will descrribe the stages incorporated in data acquisition and present the developed model representations and algorithms. Additionally we show some user interface aspects and related design decisions illustrated on our 3D modeling system.

The traditional goal of photogrammetry is to understand 3D scenes from image data. This research primarily targets real-time applications, thus human interaction during the modeling process should be avoided (Zach et al., 2003). In contrast our research goal is the creation of 3D models for visualization and computer animation. The requirements for this problem differ and human interaction during the modeling is acceptable. However, artifacts and reconstruction errors are unacceptable, because the created 3D model will be viewed and explored for example by humans in an virtual environment.

Consequently we focus in our approach on two different aspects. One aspect is related to the segmentation and interpretation of a scene. Figure 2 illustrates one side of the Josefsplatz in Vienna generated with a state of the art area based modeling algorithm. Obviously a fully automatic segmentation and interpretation of such a scene with current segmentation algorithms is not yet robust enough. In general a full segmentation procedure includes the classification of facades, roofs and any other appropriate objects. In our opinion integrating a human operator in the reconstruction loop will improve the ability to reconstruct the geometry of arbitrary cultural scenes. In fact we combine in our approach the user's intuitive image understanding and the computer's processing power.

Figure 2: Dense point cloud of one side of the Josefsplatz in Vienna. As a consequence of the outliers (sky, ground) in the created reconstruction a pre-segmentation of the scene in meaningful units like facades and roofs is necessary.

The other aspect comprises the question how the user interface is able to support the modeling process thus that the resulting 3D model is more accurate and free of disturbing outliers and reconstruction errors. Several authors (Schneiderman, 1998) discussed the direct relation between user

interfaces and the usability of a given application. However we developed our user interface having in mind the fact that humans are clumsy 3D operators, hence the user concentrates on the 2D segmentation and interpretation whereas our modeling system is responsible for the corresponding 3D information.

Figure 3 is a graphical overview of our proposed workflow. The grey shaded area will be discussed in the following sections of the paper. The workflow can be roughly seen as the composition of the following consecutive subtasks:

1. An automatic orientation procedure to obtain the relative orientation of the image sequence. This task, which is not the topic of the paper, is based on work described by Nister (Nister, 2003) and Klaus et al. (Klaus et al., 2002).

2. A highly automated feature extraction method with minimal human interaction.

3. Segmentation and classification of the captured cultural heritage supported by an intelligent user interface.

4. An automatic surface reconstruction process, which results in a coarse as well as a detailed 3D model of the scene. Furthermore this step incorporates a multi view texturing method as proposed by Bornik (Bornik et al., 2002).

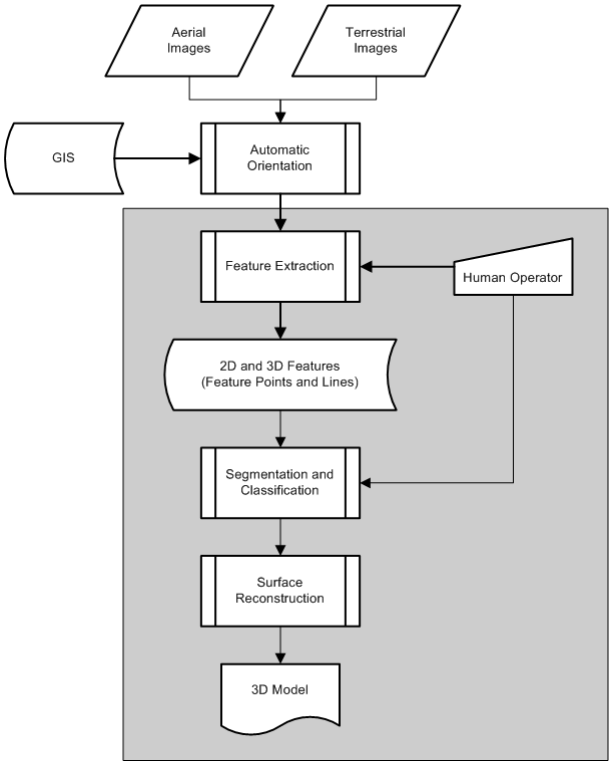


Figure 3: Sketch of the workflow. The grey shaded areas are discussed in this paper.

3.1 Data Capturing

For capturing the terrestrial images of the cultural heritage we use a calibrated high quality digital consumer camera with a 11.4 megapixels CMOS sensor. The actual capturing process consists of taking hand-held pictures with short baselines resulting in high overlap.

3.1.1 Camera Calibration The process of camera calibration is a well studied problem in photogrammetry and determines the internal parameters of a camera, which are: focal length, the position of the principal point and lens distortions. After the user has taken images of a planar calibration target from arbitrary, different positions an optimized calibration algorithm can be performed fully automatic. The method used in our 3D modeling system is described by Heikkilä (Heikkilä, 2000). In spite of recent progress utilizing uncalibrated views (Pollefeys et al., 1999) for 3D reconstruction we decided that camera calibration is a straightforward process which simplifies the reconstruction algorithm.

3.2 User Interface

In this section we illustrate the basic ideas and abilities of our implemented user interface. So far there are two basic challenges in the fields of photogrammetry: the first one is the correspondence problem and the second one is the recognition problem. In our 3D modeling system the reconstruction problem is solved by highly redundant information about the scene, particularly we utilize image sequences. Since the reconstruction problem is already solved we can focus on the recognition problem, which is handed over to a human operator, who is supported by an intelligent user interface.

As mentioned before the basic idea behind our user interface relies on the fact that humans are not good at simultaneously controlling multiple degrees of freedom. This implicates that they are not accurate 3D operators, especially with a 2D interface. In contrast computers are not limited to two eyes, thus they can handle multiple views and multiple degrees of freedom simultaneously. Therefore computers are the much better 3D operators.

Basically we combine simple 2D segmentation and interpretation tasks with more or less complex 3D modeling algorithms. Additionally we obtain a full interpretation of the selected scene in meaningful units like facades, windows or roofs. Such a system is also known as a so called monocular 3D modeling system.

Figure 4 shows the main parts of our developed user interface. Our user interface consists of an image viewer for 2D interaction and a model viewer to verify the reconstruction correctness. Furthermore a magnifier assists the human operator during the modeling process and improves the accuracy of the final 3D model. The selection of the appropriate images for the reconstruction process is accomplished by an image preview box at the bottom of the graphical user interface.

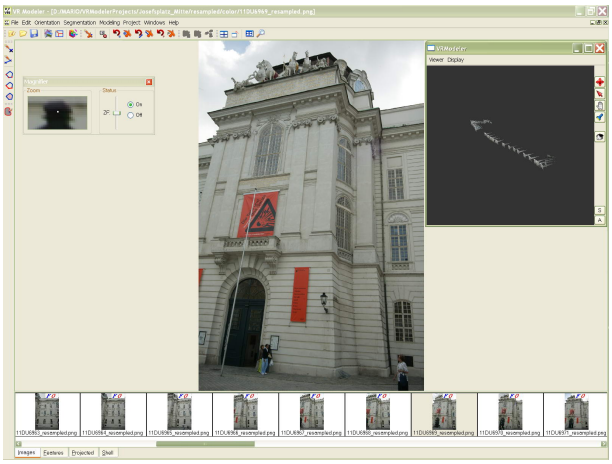


Figure 4: Illustration of our developed user interface, which includes a magnifier, an image viewer, a 3D viewer and an image preview box.

3.3 Model Representations

The intention behind our modeling system is to represent an architectural building as a set of different types of simple geometric features. Furthermore the underlying representation supports a direct relation between each 2D geometric feature and their 3D counterparts. A more detailed description of various types of features can be found in (Bauer et al., 2002). Consequently our approach deals with two types of geometric features: feature points and feature lines. The rest of the section gives a detailed description of these model representations.

3.3.1 Feature Points We propose two different approaches for the extraction of feature points: an interactive mode and a fully automatic mode. Both procedures are related to the well studied correspondence problem, which is the basic problem in stereo photogrammetry. It is the process of finding corresponding or homologues points in two or more images which are projections of the same 3D point. In fact, from corresponding points the relative orientation is estimated and additional 3D points are extracted.

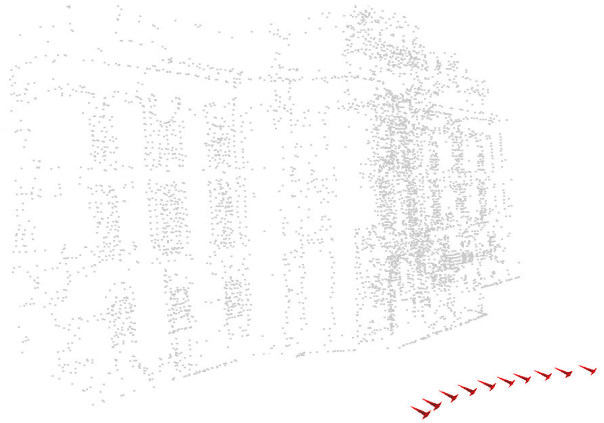


Figure 5: Automatic extracted 3D feature points and obtained camera positions from the Josefsplatz.

The automatic mode can be separated into two fundamental tasks. The first task consists of a state of the art point-of-interest detector described by Harris and Stephens (Harris and Stephens, 1988). In the second task an automatic matching procedure proposed by (Brown et al., 2003) is initiated.

In the interactive mode a human operator adds corresponding points in two images whereas the modeling system calculates the 3D points and reproject the extracted feature points into all images within the image sequence. The user is assisted by a simple point and click interface and a magnifier to obtain subpixel accuracy during the reconstruction process.

3.3.2 Feature Lines The extraction of feature lines from image sequences is a crucial step in our modeling system. However, feature lines provide a more accurate geometric information of the captured objects than feature points. Similar to feature points we distinguish between an interactive and an automatic procedure.

The automatic procedure is based on grouping of edge information, hence the results are directly related to the previous edge extraction method. Our method is based on the edge extraction and edge linking method proposed by Canny (Canny, 1986) and refined by Rothwell et al. (Rothwell et al., 1995). These algorithms extract contour chains with subpixel accuracy. The final line segments are detected by a robust estimator called RANSAC (Fischler and Bolles, 1981) and represent the input for the automatic line matching algorithm introduced by Schmid and Zisserman (Schmid and Zisserman, 2000). The purpose of the interactive procedure is to increase the set of extracted 3D lines.

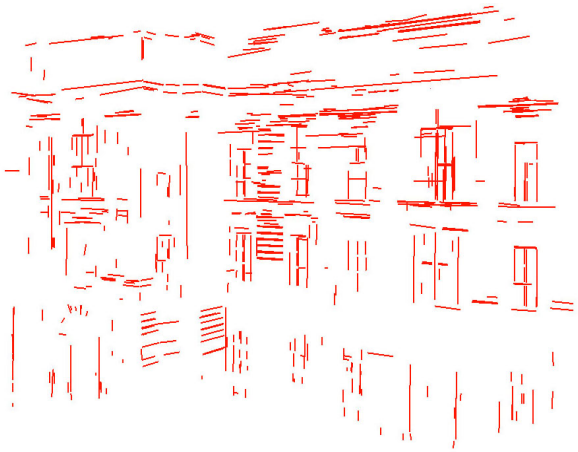


Figure 6: Overview of the extracted 3D feature line set from the Josefsplatz.

3.4 Surface Reconstruction

So far we have described the extraction of low level image features, such as feature points and feature lines from oriented image sequences. This section is dedicated to explain how these feature types are utilized to create coarse as well as detailed 3D models from the captured scene.

The overall idea of our modeling strategy is based on the fact that we combine the extracted feature types with an area based reconstruction algorithm. Actually we obtain a segmentation and interpretation of the scene into several primitives like roofs or facades.

The surface reconstruction process consists of two consecutive parts. In the first part the image features are grouped around interesting areas of the scene to obtain a coarse model of the scene. Such an initial model is represented as a set of planar object regions. Note, that the whole 2D grouping process is supported by a human operator in terms of his segmentation and interpretation abilities. Due to the direct relation between the 2D and 3D model representation the 3D surface can be easily generated. Consequently the triangulation of the planar regions and the 3D surface generation including the texturing is performed by our modeling system. This intermediate reconstruction result represents the coarse 3D model of the scene, but in general an architectural building will have additional geometric details which are not covered by a set of planar regions.

Therefore the second part of the algorithm utilizes these extracted planar object regions as an initial reconstruction plane. To recover the geometric details we focus on an iterative and hierarchical dense matching approach by exploiting the already known epipolar geometry between the images. Obviously the high detailed reconstruction is performed exclusively inside of the emphasized planar region. For every sampling point the similarity is quantified by some suitable cost function and a weighting term to allow smooth surfaces in textureless regions. This hierarchical approach converges fast and avoids ambiguities with repetitive patterns, especially often encountered in architectural buildings. The texture information of the reconstructed surface is calculated using all images and is mainly based on a multi view texturing approach described in more detail by Bornik et al. (Bornik et al., 2002).

3.5 Model Completion

Up to know we have discussed the reconstruction of a historic city centre only from terrestrial images. However to obtain a complete and consistent model of the scene it is necessary to integrate aerial images into the 3D modeling procedure.

Basically the model completion needs some human interaction and works as follows. Due to high memory demands of aerial images a human operator crops the relevant region from aerial images to complete the 3D model. Figure 7 illustrates the graphical user interface to accomplish this task.

The upgrade to a georeferenced orientation requires at least three well distributed control points. Therefore the control points taken from the aerial images are manually linked to terrestrial control points in two images. By exploiting this direct relation we calculate a transformation matrix to perform the georeferenced upgrade for the whole image sequence.

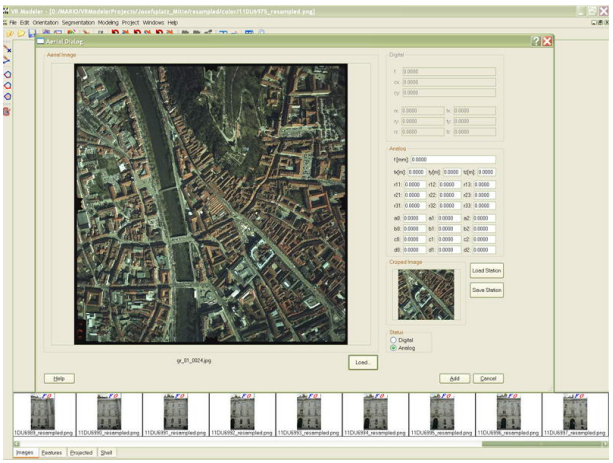


Figure 7: User Interface to add aerial images to the modeling process.

4 VIRTUAL EXPLORATION OF THE CULTURAL HERITAGE

As soon as the models are reconstructed, they can be observed in an immersive way using a virtual reality (VR) environment. The main benefits of such an environment are that the real 3D user interaction is decoupled from standard 2D input devices like mouse or keyboard and an improved visual perception due to a stereoscopic view. Interacting with 3D models using mouse or keyboard is often not sufficient or cumbersome because only restricted movements are possible. However, in a VR environment a tracked wand or pencil enables the user to intuitively navigate through or manipulate different scenes.

4.1 VR Setup

Our exploration application supports both, a stereoscopic projection wall (using shutterglasses) and a head-mounted display (HMD). The advantage of a projection wall is that more spectators can observe the scene at one time, whereas by using a HMD only one person can interact with the scene. However, the benefit of the latter method is that the user is more immersed into the scene.

As we are using the *Studierstube* environment (Schmalstieg et al., 1996) both output devices are supported transparently. For interaction, a pencil and a transparent panel are used which both allow six degrees of freedom. An optical tracking system captures position and orientation of all input devices. According to the selected setup, either the HMD or the shutterglasses are tracked by the system in order to render the correct image for the interacting person. Figure 8 shows an example of interacting with one side of the reconstructed Josefsplatz in Vienna.

4.2 Interaction

The VR application allows several interaction modes. Generally, the navigation is carried out by using the panel and the pencil. Virtual 3D buttons are displayed on top of the panel which trigger different actions (e.g. load new model,

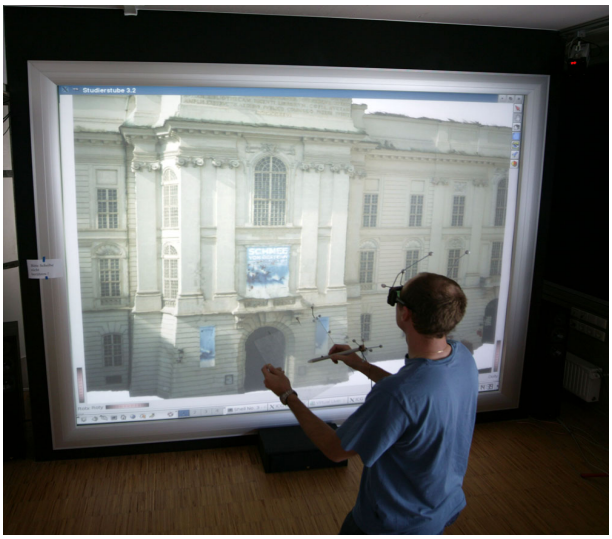


Figure 8: Interaction with a reconstructed model.

store scene, ...). Additionally, other interaction widgets are provided which allow for more complex interaction. Each single object in the scene is identified by a unique object identity. Therefore, distinct object rotation, scaling, or translation can be performed using the pencil.

If the user desires walk through a whole city, a virtual 2D map is displayed on the panel. By selecting target positions on this map, a path is calculated and the user is seamlessly guided to this location. These examples only show some ideas of exploring reconstructed cities in VR environments. However, one can imagine that far more interaction possibilities exist, which can easily be developed for such systems.

5 RESULTS

Figure 9 shows different views of the coarse 3D model of the Josefsplatz in Vienna. Note, that almost all architectural models, except statues can be reconstructed by an alignment of planar object regions.

In contrast Figure 10 illustrates a dense reconstruction of one side of the Josefsplatz. As expected the high detailed model is free of disturbing outliers. Figure 11 shows some close-ups of the detailed reconstruction to emphasize the high geometric resolution.

6 DISCUSSION AND FUTURE WORK

We have developed a new approach to reconstruct arbitrary 3D scenes, especially historic city centres from terrestrial image sequences and aerial images, by exploiting human image understanding together with computational processing power. Additionally we have outlined the implemented model representations and our user interface which is based on a monocular 3D modeling system. Basically this framework can produce 3D models of large environments more effective and accurate than current area or feature based modeling approaches. In fact the modeling process combines area and feature based modeling techniques to recover consistent and accurate models of the scene.

Beside a geo-referenced orientation of the extracted 3D model, we obtain a segmentation as well as an interpretation of the scene. Due to the complexity of historic city centres, building 3D models is time consuming and usually involving much manual effort. With our modeling system we focus on minimizing human interaction to obtain a nearly automatic reconstruction. Additionally we have sketched a method to visualize the reconstructed city centres in a virtual reality environment.

Though the results are very promising, there are several improvements that can made to our approach. First we will concentrate on extending our modeling approach to handle more complex structures. In particular we will focus on the reconstruction of statues by involving shape-from-silhouette methods into our workflow. Further we will concentrate on the integration of recognition aspects into our 3D modeling approach to further minimize human interaction and to improve the created 3D models. Additionally we will continue on improving the usability of the user interface and the performance of the developed algorithms.

7 ACKNOWLEDGMENTS

This work is partly funded by the VRVis Research Center, Graz and Vienna/Austria (`http://www.vrvis.at`). We would also like to thank the Vienna Science and Technology Fund (WWTF) for supporting our work in the the *Creative Histories - The Josefsplatz Experience* project.

REFERENCES

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A., 1999. Automatic line matching and 3d reconstruction of buildings from multiple views. In: ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5, pp. 69–80.

Bauer, J., Klaus, A., Karner, K., Schindler, K. and Zach, C., 2002. MetropoGIS: A feature based city modeling system. In: ISPRS Journal of Photogrammetric Computer Vision, ISPRS-Commission III (PCV), pp. Part B 22–28.

Bornik, A., Karner, K., Bauer, J., Leberl, F. and Mayer, H., 2002. High-quality texture reconstruction from multiple views. Journal of Visualization and Computer Animation 12(5), pp. 263–276.

Brown, M. Z., Burschka, D. and Hager, G. D., 2003. Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(8), pp. 993–1008.

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6), pp. 679–698.

(a) Left View (b) Front View (c) Right View

Figure 9: Novel views (a-c) of the Josefsplatz in Vienna represented as coarse model.

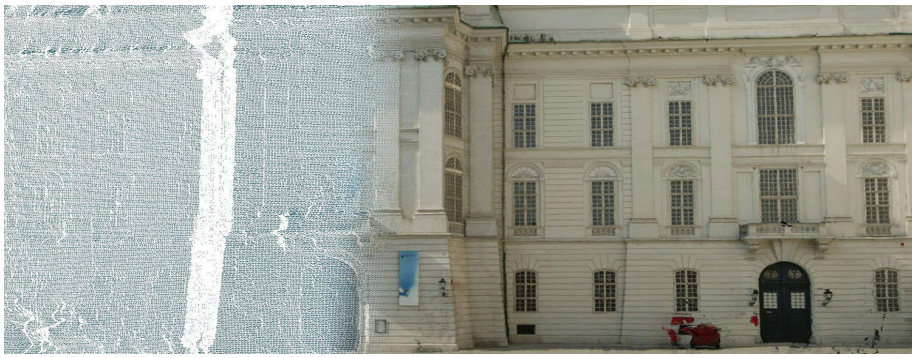
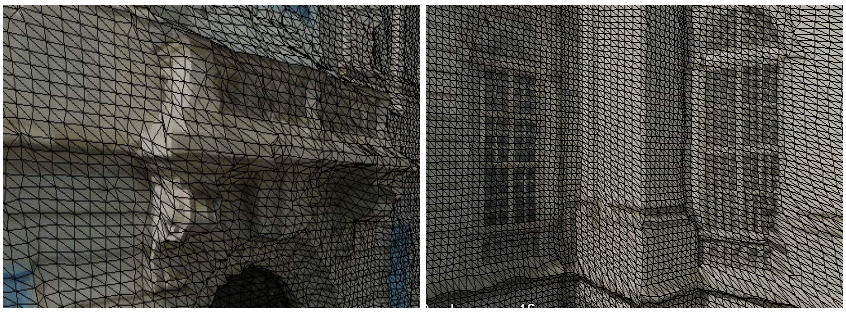


Figure 10: Dense reconstruction of one side of the Josefsplatz. To emphasize the geometric details only a part of the mesh is textured.

(a) Close-up of the Josefsplatz without overlayed wireframe to illustrate the geometric as well as texture quality.



(b) Close-up of the balcony (c) Close-up of two reconstructed windows.

Figure 11: Different close-up views illustrating the main facade. Note the high geometric resolution around the balcony (b) and the two windows (c).

Debevec, P. E., 1996. Modeling and Rendering Architecture from Photographs. PhD thesis, University of California at Berkeley, Computer Science Division, Berkeley CA.

Fischler, M. and Bolles, R., 1981. Random sample cons pp. 227-234ensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the Association for Computing Machinery 24(6), pp. 381–395.

Harris, C. and Stephens, M., 1988. A combined corner and edge detector. In: 4th Alvey Vision Conference, pp. 147–151.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Campridge University Press. ISBN: 0521540518.

Heikkilä, J., 2000. Geometric camera calibration using cicular control points. Pattern Analysis and Mashine Intelligence 22(10), pp. 1066–1077.

Horn, B., 1990. Relative orientation. International Journal of Computer Vision 4(1), pp. 59–78.

ImageModeler, 2004. http://www.realviz.com/. REALVIZ.

Klaus, A., Bauer, J. and Karner, K., 2002. MetropoGIS: A semi-automatic city documentation system. In: ISPRS Journal of Photogrammetric Computer Vision, ISPRS-Commission III (PCV), pp. Part A 187–192.

Nister, D., 2003. An efficient solution to the five-point relative pose problem. In: Computer Vision and Pattern Recognition (CVPR), pp. II: 195–202.

Pollefeys, M., Koch, R. and Gool, L. V., 1999. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. International Journal of Computer Vision 32(1), pp. 7–25.

Pollefeys, M., Koch, R., Vergauwen, M., Deknuydt, A. A. and Gool, L. J. V., 2000. Three-dimensional scene reconstruction from images. In: B. D. Corner and H. Nurre, Joseph (eds), Conference on Three-Dimensional Image Capture and Applications II, SPIE, Bellingham, Washington, pp. 215–226.

Rothwell, C. A., Mundy, J., Hoffman, W. and Nguyen, V., 1995. Driving vision by topology. In: IEEE Symposium on Computer Vision, pp. 395–400.

Scharstein, D. and Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision 47(1), pp. 7–42.

Schmalstieg, D., Fuhrmann, A., Szalavari, Z. and Gervautz, M., 1996. Studierstube - an environment for collaboration in augmented reality. In: Proceedings of Collaborative Virtual Environments, pp. 19–20.

Schmid, C. and Zisserman, A., 2000. The geometry and matching of lines and curves over multiple views. International Journal of Computer Vision 40(3), pp. 199–233.

Schneiderman, B., 1998. Designing the User Interface: Strategies for Effective Human-Computer-Interaction. Addison-Wesley. ISBN: 0201694972.

Shum, H., Han, M. and Szeliski, R., 1998. Interactive construction of 3d models from panoramic mosaics. In: Computer Vision and Pattern Recognition (CVPR), pp. 427–433.

Strecha, C., Tuytelaars, T. and Gool, L. V., 2003. Dense matching of multiple wide-baseline views. In: International Conference on Computer Vision (ICCV), pp. 1194–1201.

Zach, C., Klaus, A. and Karner, K., 2003. Accurate dense stereo reconstruction using 3d graphics hardware. In: Eurographics.