

EXPLORING RECONSTRUCTED 3D MODELS IN A VIRTUAL REALITY ENVIRONMENT

Bernhard Reitinger^a, Mario Sormann^b, Konrad Karner^b

^a Institute for Computer Graphics and Vision, Graz University of Technology,
Inffeldgasse 16/II, A-8010 Graz, Austria

^b VRVis Research Center, Inffeldgasse 16/II, A-8010 Graz, Austria
breiting@icg.tu-graz.ac.at

KEY WORDS: applications, city, exploration, interactive, real-time, reconstruction, virtual reality

ABSTRACT

This paper proposes a new way of exploring reconstructed models in 3D within a Virtual Reality environment. Different to traditional 2D desktop-based systems, 3D interaction allows for easy and intuitive navigation and selection/manipulation of these models. By using different Virtual Reality output devices, the user is able to explore immersively reconstructed 3D models in full size. The immersion is enhanced by using a stereoscopic setup. Interaction in such environments is eased by using real 3D input devices which can be used for either navigation or object manipulation. The models which can be explored in our environment are generated in a prior 3D reconstruction process where user interaction is kept at a minimum. The data exchange is seamlessly and uses VRML as data format. The main interaction is done in our application where the user can interact with different kinds of reconstructed objects. This paper presents the environment which can be used for such explorations, and necessary interaction aspects.

1 INTRODUCTION

Visualization of reconstructed cities or models in general is a necessary task after reconstruction either for exploration or for visual evaluation of the result. Often multiple different models are reconstructed which are not inherently clued together. Therefore, user interaction is necessary to combine different reconstructed models. The usual way to do that is to use a desktop-based system with mouse and keyboard, and try to perform 3D manipulation (rotation, scaling, translation). However, this can get a hard and tedious task if performed multiple times. The reason is because of the fact that the mouse only allows a 2D navigation and the desktop only displays in 2D. Even by exploring large-scale datasets (e.g. large reconstructed city parts) an immersive exploration is often desired. For instance, one scenario can be in museums where visitors can navigate through whole reconstructed ancient cities in a Virtual Reality environment. Another example is the art design of an architect combining different existing buildings with new designed models. In this case, the existing buildings must be reconstructed by a semi-automated process and in turn clued together within the VR environment. The benefits of using such an environment are clear. A real 3D interaction allows for rapid prototyping which is beneficial in the design phase.

Different requirements must be fulfilled in order to provide a realistic environment for VR applications in general and for exploring reconstructed 3D models specifically:

- **Real-time rendering:** For an interactive environment frame-rates beyond 30 frames/second are required.
- **Realism:** However, realism is important for users in order to get an improved immersion.
- **Easy interaction:** The almost most important issue of such environments is the human-computer interface

where interaction devices and interaction metaphors are equally important.

- **Stereoscopy:** An important issue of an VR system is stereoscopy which enhances the realism and immersion.

This paper will demonstrate that these four issues are addressed by our system. The proposed application is mainly based on previous work in the field of AR/VR and virtual liver surgery planning (Bornik et al., 2003) which is an on-going research project at the Graz University of Technology in cooperation with the Departments of Radiology and General Surgery at the University Hospital Graz and the Department of Electrical and Computer Engineering of the University of Iowa.

The idea and motivation for the hereafter presented work is given by merging new algorithms for reconstructing the city center of Graz (which was awarded World Cultural Heritage by UNESCO in 1999) with interaction techniques in a Virtual Reality environment. The buildings of the inner city area are generated by using terrestrial images of facades taken by a hand-held digital consumer camera. Details of this algorithms can be found in Section 3.2. These reconstructed models are then imported into the VR environment by using the VRML standard and combined with 3D interaction elements.

The rest of this paper is organized as follows: The next section presents some related work in the field of exploration and rendering of 3D models. Section 3 describes the environment itself and outlines the user interface and some interaction elements of the application. In Section 4, some results and screenshots are shown presenting a manifold environment. Finally, Section 5 and 6 conclude with a discussion and some possible future work.

2 RELATED WORK

Using Virtual Reality technologies for exploring 3D models is not a novel research area. A lot of work has been done in this field focusing on rendering, computer graphics methods and so forth. However, user interaction for exploration and easy user handling has not got much attention. Traditional VR applications can be found in various fields: medical (Bornik et al., 2003, Stansfield et al., 2000), education (Kaufmann et al., 2000, Dede et al., 2000), entertainment (Brill, 2001), and arts (Bernardini et al., 2002) to name only a few.

As far as cultural heritage is concerned, some research work was published recently. Grabner et al. presented a combination of the MURALE project and VR technologies in order to explore archaeological sites in space with simple user interaction (Cosmas et al., 2001) (Grabner et al., 2002). Another project called “OpenView” was presented which also utilizes VR technologies to visualize archaeological 3D models (Sechidis et al., 2004). However, the focus of this project is also not interaction but visualization and scripting.

Apart from VR interaction but focusing on real-time rendering, Zach et al. presented some work for modeling and visualizing cultural heritage datasets of Graz (Zach et al., 2001). By using a database management system, an intelligent level of detail concept was developed which guarantees fast rendering.

In the field of user interaction, Bowman et al. gave a broad overview of 3D interaction and user interface in (Bowman et al., 2001). They proposed to categorize user interaction in three different groups of tasks:

- *Navigation*
- *Selection/manipulation*
- *System control*

For large-scale 3D environments, navigation is the most prevalent action in order to perform comfortable movements between two distant locations. Selection and manipulation accomplish at least one of these tasks: object selection, object positioning, object rotation. Various techniques have been suggested for this kind of interaction, one example is the Go-Go technique proposed by Poupyrev et al. where the user’s reach is extended by a non-linear mapping applied to the user’s hand (Poupyrev et al., 1996).

The first two interaction modes (*navigation* and *selection / manipulation*) are also the most prevalent used within our environment. The first one is used for exploring large-scale models, whereas the latter is target for modeling and design.

3 TOOLS AND METHODS

After presenting some related work, this section describes the used and developed methods for exploring reconstructed 3D models in a VR environment.

3.1 The Virtual Reality Environment

While there is no common definition of the term *Virtual Reality*, it can be best described by three features:

Immersion-Interaction-Imagination

Immersion can be gained by using a tracked stereoscopic device like a head-mounted display (HMD) and by rendering stereo images. The immersed user should be able to *interact* with the virtual reality as it would be real. And finally, the *imagination* refers to the user’s mind capacity to perceive nonexistent things (Burdea and Coiffet, 2003).

3.1.1 VR Setup In order to develop a VR application, the utilized hardware for input/output plays an important role. Our developed system is running on the setup installed at the Institute for Computer Graphics and Vision, Graz University of Technology and mainly consists of the following components:

Output

- A stereoscopic **large-screen back-projection wall** consisting of an active stereo DLP projector (Barco Galaxy) displaying stereo images at a resolution of 1280×1024 pixels 120 times a second (60 frames for each eye). The projector is driven by using an off-the-shelf graphics hardware which must support quad-buffer output. As the system is active stereo, shutter-glasses must be used which are synchronized with the video stream projected on the wall.
- Alternatively, a **HMD** can be used which provides an improved immersion if the field-of-view is large enough.



Figure 1: Tracked head-mounted display.

Input

- The most important input devices in our environment are a tracked **pencil and panel**. The panel is a transparent Plexiglas plate on which virtual steering widgets can be displayed. By using both in combination, a real 6 degrees-of-freedom (DOF) device is provided which can be used for any type of interaction.

- Additionally a tracked **turntable** can be used as interaction device for rotating objects in a very intuitive way. Only a one-axis rotation is supported by this device.

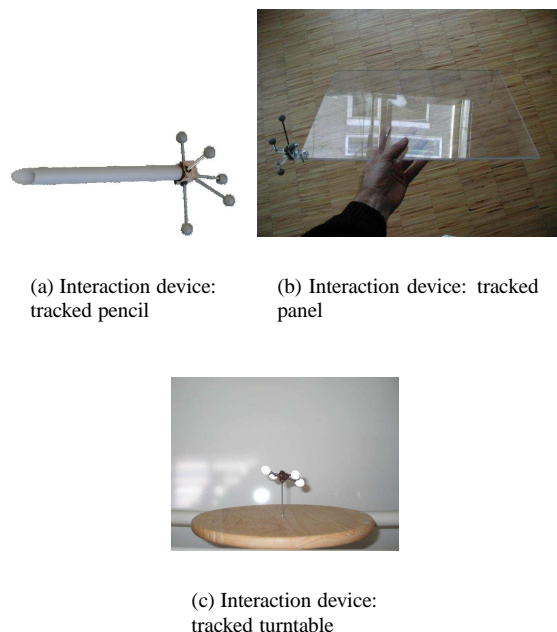


Figure 2: Input devices with 6 DOF.

Tracking In order to get the correct position and orientation of both, the input and output devices, an optical tracking system by A.R.T. equipped with four cameras is used. In order to track one devices, at least two cameras must “see” the targets. By using four cameras, the robustness of this system is improved.

3.1.2 VR Software The basic library which is used in our environment is called the “Studierstube” which is an on-going project initiated in 1996 for providing low-level methods in Virtual as well as in Augmented Reality environments (Schmalstieg et al., 1996). The main interaction devices are - as previous stated - a pencil and a personal interaction panel (PIP) (Szalavari and Gervautz, 1997).

The PIP in combination with a pencil can be used for all kind of interaction like system control, navigation, or object manipulation/selection. By projecting 2.5D virtual widgets on the translucent panel, menu items can be selected utilizing the button mounted on the pencil. The main benefit of this type of interaction is the intrinsic contained force feedback which helps the user to accomplish the widget selection. However, if application steering (system control) gets very complex, this type of interaction has its disadvantages because widgets might get very small. Therefore, a one-handed interaction metaphor is desired which in addition does not lead to hand and arm strain if used for a longer period of time. The *Virtual Cockpit* is one possibility to overcome these physical strains (Terbu, 2004) which is also an integrated part of the “Studierstube”.

The core of the “Studierstube” is based on the scene-graph library called *Coin* (SIM, 2004) which handles correct and

fast rendering transparently. Moreover, the scene-graph based metaphor allows us to support the interaction with multiple reconstructed models concurrently in one scene.

3.2 Reconstruction

Before focusing on the user interface and interaction elements of our VR environment, the developed methods for the reconstruction of 3D models are presented. The reconstruction of 3D models from image sequences is an active field within the photogrammetric and computer vision community (Debevec, 1996), (Pollefeys et al., 2000) and (Schmid and Zisserman, 2000). In the following few paragraphs we outline the consecutive steps of the semi-automatic reconstruction process. In fact the work-flow consists of the following tasks:

1. Data capturing and preprocessing
2. Automatic orientation
3. Feature extraction
4. Segmentation and classification
5. Surface reconstruction

The actual recording consists of taking hand-held pictures of the object with a calibrated digital consumer camera with short baselines. The camera is calibrated by taking images of an indoor calibration target from arbitrary viewpoints and performing an automatic calibration algorithm as proposed by (Heikkilä, 2000).

One of the basic problems in the field of photogrammetry is the correspondence problem, which is to identify 2D points in two images that are projections of the same 3D point in the world. From corresponding points within the image sequences the relative orientation and the 3D positions of the corresponding points can be estimated, as illustrated in Figure 3.

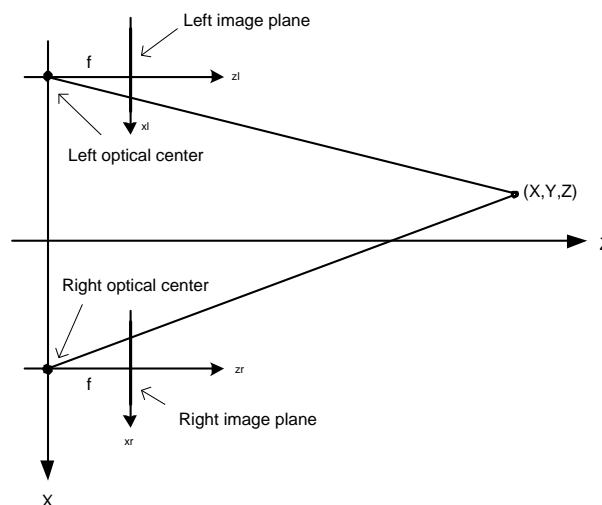


Figure 3: Illustration of the basic correspondence problem.

A reliable calculation of the relative orientation is based on an accurate point of interest (POI) extraction proposed by Harris (Harris and Stephens, 1988) followed by an affine invariant area based matching approach. Since the automatically matched correspondences can still contain few outliers a robust estimation method called RANSAC (Fischler and Bolles, 1981) is utilized. In order to obtain the orientation of the whole image sequence it is necessary to determine the scale factor. This task is accomplished by utilizing corresponding points in at last three images. A first result of an automatically oriented image sequences is illustrated in Figure 4.

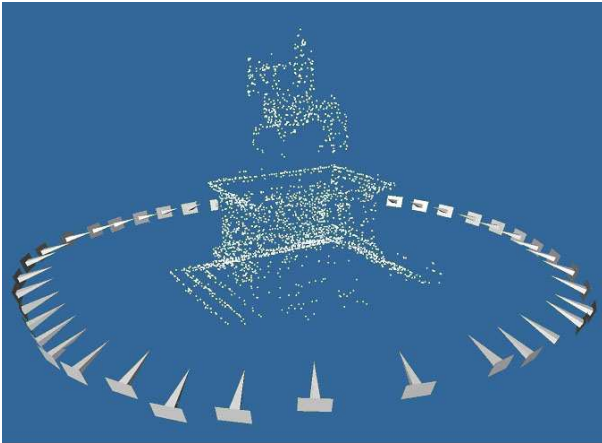


Figure 4: Corresponding 3D points which are obtained during the automatic orientation procedure.

Once we have determined the orientation of the image sequence we are able to extract different types of model representations from the photographs automatically by employing different feature-based matching techniques. Our reconstruction approach deals with two different types of simple geometric primitives: feature points and feature lines. These different geometric primitives are further combined with an area based reconstruction algorithm. Essentially the whole 2D selection and grouping process is supported by a human operator in terms of his segmentation and interpretation abilities. Such a segmentation is important to reduce the outlier rate in the following dense matching procedure and to obtain an interpretation of the scene in meaningful units like facades and roofs.

To obtain all geometric details of an architectural building we concentrate on an iterative and hierarchical dense matching approach. For every sampling point the matching procedure optimizes a cost function, which contains the similarity between the template windows and a regularization term to favor smooth surfaces in textureless regions. A reconstruction result illustrating a statue in front of the *Landhaus* building in Graz is shown in Figure 5.

3.3 Model Management

Each reconstructed model is represented by an *Entity* in the scene-graph (see Figure 6). An *Entity* consists of a *Transform* node which represents the local coordinate transformation of each single model (rotation / translation

/ scale) and is identified by a unique ID. Each *Entity* is localized by using a bounding box operation on the scene-graph. If the pencil is inside the model's bounding box, a wireframe is drawn to give a visual feedback indicating that the selected operation can be performed on this currently selected object.

The *ModelManager* which is a super-node of all *Entities* is responsible for enabling and disabling each single model. If a new model is inserted into the scene, it gets registered at the *ModelManager*.

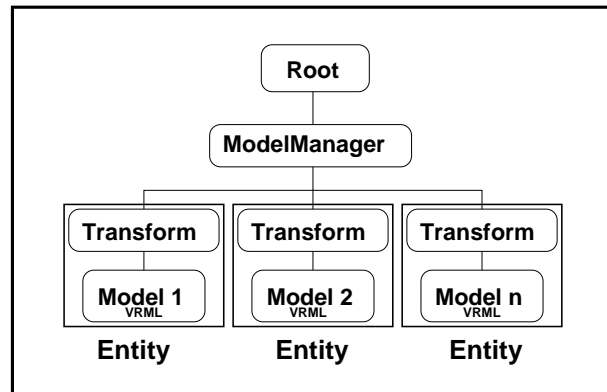


Figure 6: Scene-graph managing multiple 3D reconstructed models each having a unique *Transform* node.

3.4 User Interface/Interaction

The user interaction is carried out by using the PIP and a pencil. The complete menu for system control is projected on the personal interaction panel including different modes of interaction. Generally, we have to distinguish between operational (system control) widgets, translation (manipulation) widgets, and navigation widgets. Operational widgets provide the following functionality:

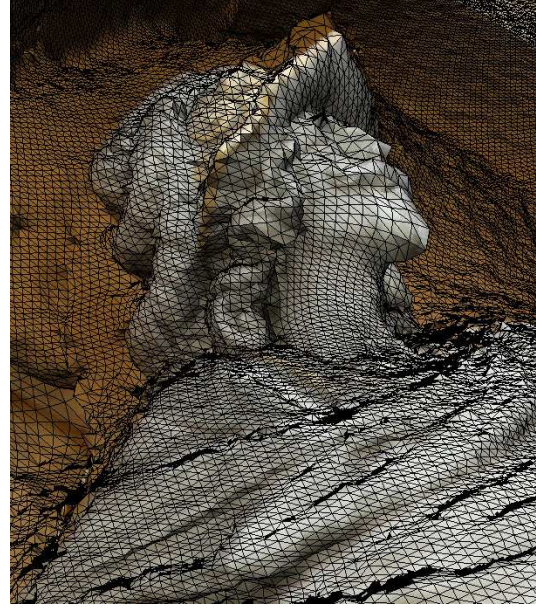
- *load/store*: load new reconstructed models, and store the complete scene-graph
- *wire/solid*: change the rendering mode of the selected model(s)
- *group/ungroup*: multiple models can be grouped to one coherent block and allows consequent group operations
- *listbox*: a listbox shows all loaded models by name

For each selected model or coherent model group, different manipulation operations are provided which are only valid if models are small in size. By using the pencil, either a rotation, translation, or a scaling can be applied to the currently selected model (group). Six DOF allow fast and intuitive model manipulation and rapid prototyping can therefore be implemented very easily.

The last type of operation provides interaction elements for navigation of large-scale models. For instance, if a whole reconstructed city part should be explored (e.g. the old town of Graz) a 2D map showing a top-view of the 3D



(a) Dense 3D surface model illustrating the geometric as well as texture quality.



(b) Close-up with overlaid wireframe.

Figure 5: Dense reconstruction and close-up of the front side of the *Landhaus* building in Graz.

model is provided and projected onto the navigation sheet of the PIP. By using again the pencil, distant locations can be managed in a click and drag fashion. A path is calculated from the current location to the target point and the user is smoothly translated on this path to the desired location.

Another interaction mode for large-scale models is *walk-through*. The user is initially positioned at a starting location, and by pointing the pencil in one direction, the user is moved smoothly into this direction. This interaction mode is very simple and effective and can easily be used by novice VR users. An virtual direction arrow indicates the current direction. The walk can be initiated and stopped by using the mounted button on the pencil.

For very complex scenarios, a pre-calculated path is the best choice replacing human's interaction. This interaction metaphor can often be seen in combination with narrating virtual tour guides. Another possibility is to provide path planning where the user can specify different sight-seeing places of choice. The CAVE (Cruz-Neira et al., 1992) where multiple users are immersed builds the best platform for these types of presentation.

4 RESULTS

This paper presented an idea of merging reconstructed models gained through semi-automated algorithms and Virtual Reality methods for exploration in an immersive environment. The frame-rates strongly depend on the number of polygons and the used textures for the facades. However, at least 30 frames per second were achieved in all tested models. The exploration application can be either used with a

large-screen projection system or by using a HMD with improved immersion. Different interaction modes are provided in order to interact with each single model or with a group of selected objects.

Figures 7, 8, 9 show different screenshots taken by a digital camera. As output device the stereoscopic back-projection wall was used. Figure 7 displays two famous reconstructed towers of Graz. Figure 8 visualizes the "Landhaus" statue which is a very complex model reconstructed in 3D with our algorithms. Finally, Figure 9 shows interacting with two different models which are positioned next to another by using the pencil as interaction device.

5 DISCUSSION

As shown in the previous section, our VR exploration application is manifold. Different reconstructed models can be visualized and investigated. By recapitulating the four requirements of a VR system stated in the introduction, all of these are fulfilled by our environment:

- **Real-time:** all loaded models can be inspected with at least 30 frames per second
- **Realism:** by using robust 3D reconstruction algorithms, the models reflect the real world very good
- **Easy interaction:** by using intuitive 6 DOF interaction devices, the application handling is comfortable
- **Stereoscopy:** by utilizing either the projection wall or the HMD, stereo images are supported and enhance the immersion different to traditional 2D visualization on desktop-based systems

However, the main problem can be seen in the real-time rendering aspect. As the reconstructed surface models can often have a high polygon resolution, the frame-rate can drop if loading multiple models concurrently. A level of detail strategy (like presented in (Zach et al., 2001)) or a surface simplification based on quadric error metrics (also see (Garland and Heckbert, 1998)) should be applied in order to get multiple resolutions of one model. Moreover, view-dependent texture mapping can improve the rendering of such texture intense buildings (also see (Sormann et al., 2003)).

6 FUTURE WORK

The presented application is not yet fully exploit. Far more interaction modes can be integrated. Another idea is to combine the interaction required for the modeling with VR technologies minimizing the period of interaction time. Hence, a full-blown application for 3D reconstruction and visualization/exploration in VR can be offered to the target users. Moreover, we have to evaluate new visualization possibilities and input devices which may be more beneficial for such kind of application.

7 ACKNOWLEDGEMENTS

This work is partly funded by the VRVis Research Center, Graz and Vienna/Austria. 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

Bernardini, F., Rushmeier, H., Martin, I., Mittleman, J. and Taubin, G., 2002. Building and digital model of michelangelo's florentine pieta. *IEEE Computer Graphics and Applications* pp. 59–67.

Bornik, A., Beichel, R., Reitingner, B., Gotschuli, G., Sorantin, E., Leberl, F. and Sonka, M., 2003. Computer aided liver surgery planning: An augmented reality approach. In: R. Galloway (ed.), *Medical Imaging 2003, Proceedings of SPIE*, Vol. 5029, SPIE Press.

Bowman, D., Kruijff, E., LaViola, J. and Poupyrev, I., 2001. An introduction to 3-d user interface design. *Presence* 10(1), pp. 96–108.

Brill, L., 2001. Illusion: Jumping feet first into immersive entertainment. *Real Time Graphics* 10(4), pp. 6–8.

Burdea, G. and Coiffet, P., 2003. *Virtual Reality Technology*. 2nd edn, Wiley-Interscience.

Cosmas, J., Itegaki, T., Green, D., Grabczewski, E., Weimer, F., Gool, L., Zalesny, A., Vanrintel, D., Grabner, M., Karner, K., Leberl, F. and Schindler, K., 2001. 3D MURALE: A multi-media system for archaeology. In: *Proceedings of ACM Siggraph conference VAST2001*.

Cruz-Neira, C., Sandin, D., DeFanti, T. and Kenyon, R., 1992. The CAVE: audio visual experience automatic virtual environment. In: *Communications of the ACM*, Vol. 35number 6, pp. 64–72.

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

Dede, C., Salzman, M., Loffin, B. and Ash, K., 2000. *Innovations in Science and Mathematics Education: Advanced Designs for Technologies of Learning*. Erlbaum, Hillsdale, chapter *The Design of Immersive Virtual Learning Environments: Fostering Depp Understanding of Complex Scientific Knowledge*, pp. 361–413.

Fischler, M. and Bolles, R., 1981. Random sample consensus: 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.

Garland, M. and Heckbert, P., 1998. Simplifying surfaces with color and texture using quadric error metrics. In: *Proceedings of the IEEE Visualization 98*, pp. 263–270.

Grabner, M., Schwann, G., Schindler, K., Sormann, M. and Hynst, S., 2002. Navigation in virtual reality cultural heritage sites. In: *Workshop on 3D Virtual Heritage*, Geneva.

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

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

Kaufmann, H., Schmalstieg, D. and Wagner, M., 2000. Construct3d: A virtual reality application for mathematics and geometry education. *Education and Information Technologies* 5(4), pp. 263–276.

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.

Poupyrev, I., Billingham, M., Weghorst, S. and Ichikawa, T., 1996. The Go-Go interaction technique: Non-linear mapping for direct manipulation in VR. In: *Proc. of the ACM Symposium on User Interface Software and Technology*, pp. 79–80.

Schmalstieg, D., Fuhrmann, A., Szalavari, Z. and Ger-vautz, 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.

Sechidis, L., Gemenetzi, D., Sylaiou, S., Patias, P. and Tsioukas, V., 2004. Openview: A free system for stereoscopic representation of 3d models or scenes. In: IAPRS, Vol. 35, pp. 819–823. ISSN 1682-1777.

SIM, 2004. Reference. <http://www.coin3d.org/>.

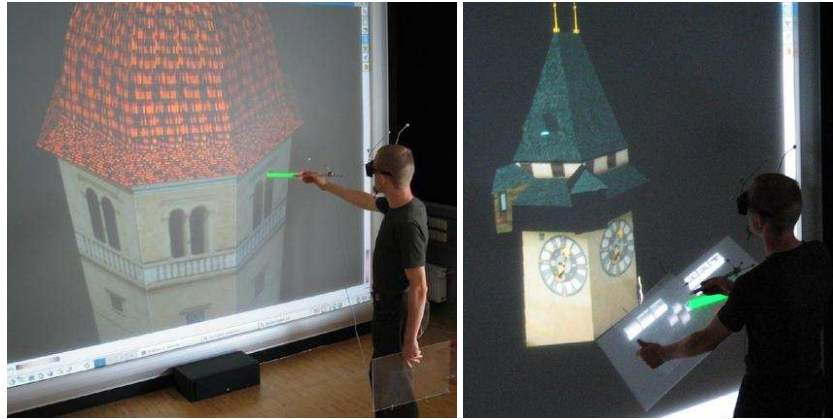
Sormann, M., Zach, C. and Karner, K., 2003. Texture mapping for view-dependent rendering. In: Proceedings of SCCG 2003, pp. 146–155.

Stansfield, S., Shawver, D., Sobel, A., Prasad, M. and Tapia, L., 2000. Design and implementation of a virtual reality system and its application to training medical first responders. Presence 9(6), pp. 524–556.

Szalavari, Z. and Gervautz, M., 1997. The personal interaction panel - a two-handed interface for augmented reality. Computer Graphics Forum 16(3), pp. 335–346.

Terbu, A., 2004. Virtual Cockpit: An Alternative Augmented Reality User Interface. In: In Proc. of the 8th Central European Seminar on Computer Graphics (CESCG) 2004, pp. 31–37.

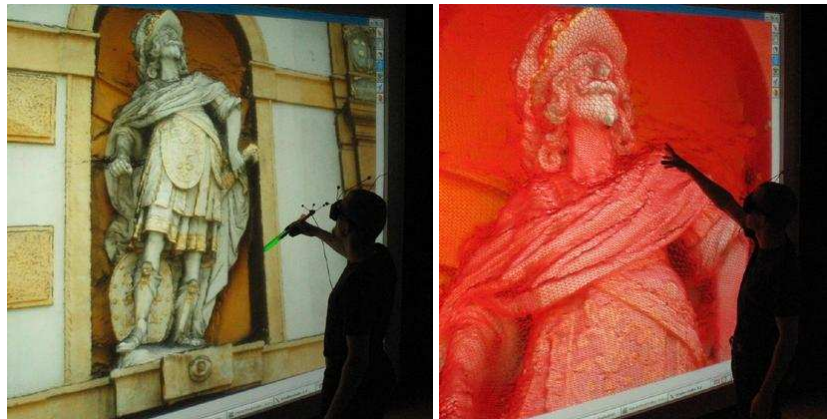
Zach, C., Klaus, A., Bauer, J., Karner, K. and Grabner, M., 2001. Modeling and visualizing the cultural heritage data set of graz. In: Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage, pp. 219–226.



(a) "Glockenturm" in Graz

(b) "Uhrturnm" at Grazer Schlossberg

Figure 7: Visualizing two famous towers of Graz ("Glockenturm" having 8570 triangles and 30 textures, and "Uhrturnm" having 7892 triangles and 51 textures).



(a) "Landhaus" statue in Graz

(b) Statue visualized in wireframe

Figure 8: Interacting with more complex "Landhaus" statue of Graz having 524288 triangles.

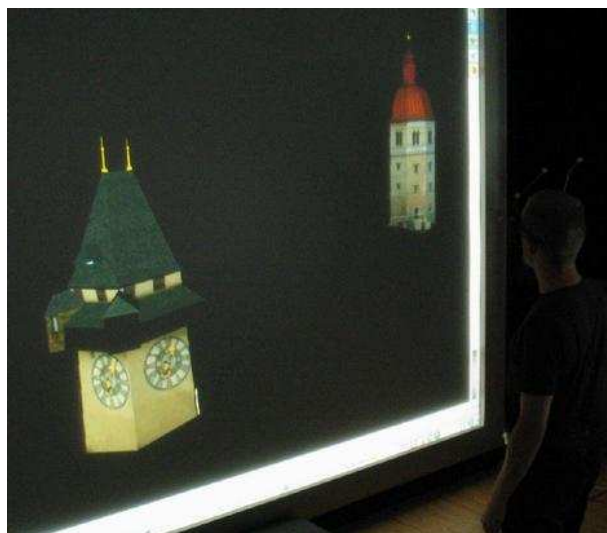


Figure 9: Two loaded reconstructed models ("Uhrturnm" and "Glockenturm") positioned side-by-side.