EPOCH Conference on Open Digital Cultural Heritage Systems (2008)
David Arnold, Franco Niccolucci, Daniel Pletinckx, Luc Van Gool (Editors)

Tools for Populating Cultural Heritage Environments with Interactive Virtual Humans

D. Arnold[1], A. Day[2], J. Glauert[2], S. Haegler[5], V. Jennings[2], B. Kevelham[3], R. Laycock[2], N. Magnenat-Thalmann[3], Jonathan Maïm[4], D. Maupu[4], G. Papagiannakis[3], D. Thalmann[4], Barbara Yersin[4] and K. Rodriguez-Echavarria[1]

[1]University of Brighton - Brighton, United Kingdom
[2]University of East Anglia - Norwich, United Kingdom
[3]UNIGE - Geneva, Switzerland
[4]EPFL - Lausanne, Switzerland
[5]ETH Zürich, Switzerland

Abstract

Modern 3D VR systems rely heavily on the interplay of heterogeneous technologies. Because of this inherently interdisciplinary character, VR domain can be viewed as a melting pot of various technologies which although complementary are non-trivial to put together. Frameworks can be used to address this challenge as they offer advantages such as reusability of components, as well as ease of replacements, extensions, and adaptations. Hence, this paper presents developments within the EPOCH project, in particular the Characterize NEWTON, to improve and release frameworks that support the incorporation of avatars in interactive real-time 3D VR systems. The purpose is to enable avatars to be interactive and to react to model metadata; thus adding realism and engaging the user's interest. This vertical middleware offers the advantage to be based on open source generic frameworks, such as OpenSceneGraph and OpenSG as well as offering complementary functionalities.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Three-Dimensional Graphics and Realism]: Virtual reality I.3.4 [Graphics Utilities]: Graphics packages

1. Introduction

Once Cultural Heritage (CH) data has been captured, stored, and interpreted; CH professionals usually turn their attention to the display of their findings in interactive 3D Virtual environments. Computer graphics provides a suitable medium to facilitate the visualisation of the CH artefacts within the appropriate context. It enables virtual worlds to illustrate how the artefact was used and interacted with. However, simply watching an animation or real time simulation is a passive method of interaction, a more immersive user experience is obtained through the active exploration of the artefacts. For this, it is not only necessary to create a 3D scene but also to add realism to it by including virtual humans, often referred to as avatars.

Modern 3D VR systems rely heavily on the interplay of heterogeneous technologies. Because of that inherently interdisciplinary structure, the VR domain can be viewed as a melting pot of various technologies which although complementary are non-trivial to put together. Many of these technologies gain a lot of individual R&D interest but still it is not generally understood nor is there barely any accepted guidelines and approaches in the matter of integration of those functional artifacts under the roof of one consistent framework. In other words, there are many masterpieces of atomic technologies but still there is a lack of a well understood and generally accepted strategies for putting them together so that they would constitute a single entity, which is larger than the sum of its parts. The missing element is an open source framework which would curb the complexity and make the resulting system machinery a consistent and seamless unity, leaving at the same time open handles and hooks for replacements and extensions.

It becomes visible that the complexity of such systems reaches the levels that cannot be handled anymore efficiently by methodologies and technologies of today. Object oriented toolkits, knowledge of well established architectural and development patterns plus human skills and experience still help and do the job but in order to stay on the cutting edge of tomorrow development speed, massive reusability of components, ease of replacements, extensions, adaptations, reconfigurations and maintenance must be addressed. These are exactly the features offered by the frameworks.

Within the context of the EPOCH project, in particular the Characterize NEWTON, we aimed *to improve and release frameworks that support the development of interactive audio-visual real-time 3D VR systems*. In particular, environments which feature real-time virtual character simulation with state-of-the-art clothing. For this, the main challenges undertaken by this research were: elaboration of methodology, guidelines, architectural, design and behavioral patterns leading to the construction of vertical middleware frameworks. This resulted in the release of an easy to use set of tools to incorporate avatars that react to model metadata adding realism and interest for users. This vertical middleware offers the advantage to be based on open source generic 3D graphic frameworks, such as OpenSceneGraph [Ope08a] and OpenSG [Ope08b] as well as offering complementary functionalities.

The following sections will describe the frameworks resulting from this work: a) the release as open source of the VHD++ kernel and plug-in and b) the improvement of the UEA Scene Assembly Toolkit. The former concerns a flexible real time framework, with the emphasis on large scale character animation, whereas the latter provides a user intuitive method for combining various elements of a cultural heritage environment into a common framework. The "EPOCH Scene Renderer" imports, in Collada format, a scene constructed using the Scene Assembler and offers a guided tour of the cultural heritage site. The navigation of the site is achieved through conversing with a virtual human via a natural language processing module capable of multiple languages. The "EPOCH Scene Renderer" is described in its application to visualising the historic reconstruction of 17th Century Wolfenbuttel in Section 4. A brief overview on the usability and acceptability for the type of application produced by this framework is also described.

2. VHDPlus

UNIGE and EPFL have been actively involved in the EPOCH project contributing with virtual human simulation technologies as part of the EPOCH Common Infrastructure in both showcases as well as the Characterize NEWTON projects by adapting and releasing their core platform vhdPLUS as open-source tool for the cultural heritage community. The vhdPLUS Development Framework is a modern, fully component oriented simulation engine and software middleware solution created by and reflecting many years of the R&D experience of both the MIRALab, University of Geneva and VRlab, EPFL labs in the domain of VR/AR and virtual character simulation [PPM*03].

vhdPLUS is a highly flexible and extensible real-time framework supporting component based development of interactive audiovisual simulation applications in the domain of VR/AR with particular focus on virtual character simulation technologies (see figure 1). It relies heavily on multiple, well established Object Oriented design patterns, uses C++ as the implementation language and Python as a scripting language.

vhdPLUS has been released as open source as part of the EPOCH-NEWTON Characterize activities. In support of this release, and as per the Common Infrastructure activities, we have created several resources that should lead to a more effective use of vhdPLUS for users in general, but especially for those involved in the Characterize activities. Some of the components in the vhdPLUS version include:

- OpenSceneGraph rendering.
- OpenSG based rendering for static objects.
- VRML97/HANIM1.1 low level parser library.
- Helper library for Virtual Human control: libvhdOSGExt
- Configuration of vhdPLUS through XML files

To allow the interaction between vhdPLUS and OpenSG a new Service (a vhdPLUS plug-in) has been written that allows for the rendering and placement of geometry through OpenSG. Furthermore a library has been provided for the loading and animation of HANIM1.1 virtual humans and we included a service template with basic code to demonstrate the inclusion of new services into vhdPLUS.

Figure 1: *Screenshots from the final 3D interactive real time virtual heritage simulations based on vhdPLUS*

A further explanation of the use of XML in combination with vhdPLUS has been given on the website: [vhd08b]. We have also included basic doxygen generated documentation showing the structure of the various components that make up vhdPLUS as well as a paper demonstrating the use of VHD++ (and therefore vhdPLUS) in cultural heritage contexts [NMTCY07] and [MTP06]. Based on various user inputs, UNIGE has updated the basic building environment and scripting tools for easier adoption of the framework. EPFL has proceeded with the addition of specialised to virtual human simulation math library.

2.1. Availability

vhdPLUS has been made available through Sourceforge [vhd08a], and has been released under the LGPL2 license. It is accompanied by a website [vhd08b] in which we have made available a number of documents [PPM*03] and [MTP06] detailing the full structure of vhdPLUS as well as its capabilities and uses. Especially [Pon04] gives a full description of VHD++ (vhdPLUS's parent framework) including information plug-in structure (called Services in vhdPLUS) as well as XML initialisation. Since 2007, more than 200 downloads have illustrated the interest of the open source community on the above framework.

3. UEA Scene Assembly Toolkit

The University of East Anglia (UEA) has been, as part of the Characterize NEWTON, involved in improving a toolkit for 3D scene creation. This toolkit supports the user through an assembly pipeline by offering the following components:

1. Terrain Converter: converts terrain data from a regular grid to a ground model.
2. Avatar Research Platform (ARP) toolkit: supports the design and

animation of Virtual Humans capable of speech and sign language.

3. Scene Assembler: used to combine objects in the scene using scriptable operations and manual tweaking. For this, every command is stored in an XML script format for rapid semiautomatic assembly of scenes.

The terrain converter component (see figure 2) of the scene assembler imports either a 3D mesh in Wavefront's OBJ format or a two dimensional image. The image can be processed automatically and a regular grid of triangles extracted. Each vertex in the grid is assigned a height value, which is calculated from the pixel in the image. Whichever approach is employed for the generation of the terrain, it may be rendered efficiently in real-time using ROAM [DWS*97], and processed automatically to derive a height map. The height map is obtained by rendering from a viewpoint perpendicular to the ground using orthographic projection. The resulting pixels rendered into the frame buffer provide a two dimensional image containing grey values, where the intensity of each grey value represents the coordinate in the y axis. This information is utilised by the ARP to ensure that the virtual humans have the potential to have their feet clamped to the terrain.

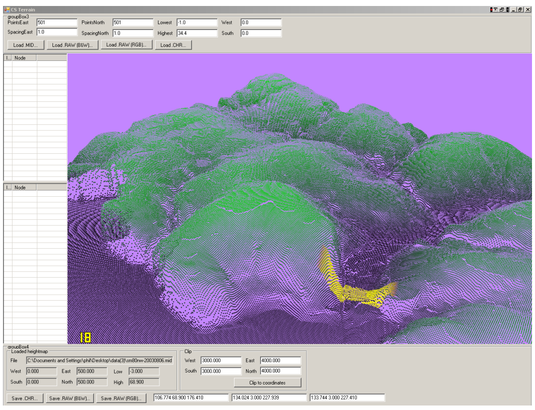

Figure 2: *Scene Assembly Toolkit: Terrain Converter component*

The Avatar Research Platform (ARP) toolkit includes tools for conventional mesh attachment and design. Some of its features include:

- Advanced tools for facial morph creation (as shown in figure 3).
- Bespoke avatar creation for procedural animation (UEA Animgen).
- Imports industry standard Maya and FBX.
- Exports to Maya and FBX.
- ARP format used by EPOCH Scene Renderer.

A propietrary extension for COLLADA [Col08] standard is used as the format for exporting from the ARP toolkit. The Scene Assembler can import virtual humans in this format into scenes. To work smoothly, it has been required to resolve some issues in OpenSG with the dynamic nature of virtual humans, such as memory overheads, rendering speeds, synchronization of audio and visual components important for speech. This was done by utilising "deprecated" interfaces that allow low-level use of OpenGL. Furthermore, interfaces have been provided for attaching hyperlinks to objects so that users can access related resources, for example websites.

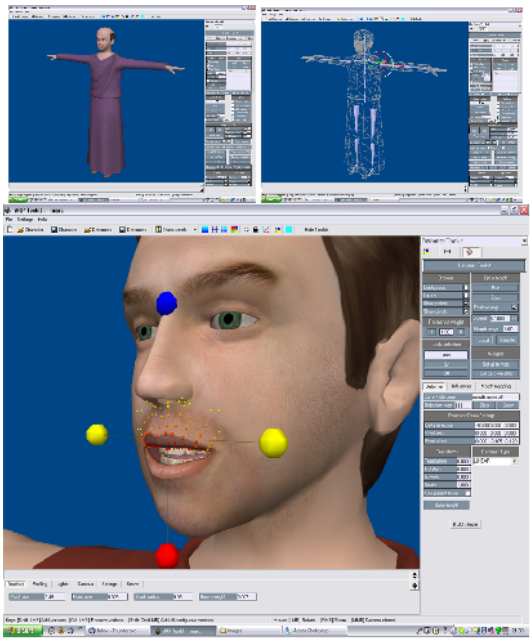

Figure 3: *Screenshots of UEA Avatar Research Platform (ARP)*

As illustrated in figure 4, the Scene Assembler tool creates a final scene by adding simple static geometry objects from popular modelling software (3DSMax, Maya, etc.). In addition, billboard objects defined by text are used to represent trees, lamp posts, and other street furniture. Objects from a group can be "planted" with a single click, especially useful for rapidly adding vegetation. Then, individual virtual humans are placed in the scene with animations as shown in figure 5. The Scene Assembler tool can export scenes in .OBJ format (static objects only) or COLLADA with extensions. Rendering of the final scene is based on the open source OpenSG rendering framework and utilizes occlusion culling techniques for rapid animation.

The occlusion culling approach requires that polygons or collections of polygons are identified in the scene to function as the occluders. During rendering any polygons that would be rendered completely behind the occluder, for a particular viewpoint, are culled. The Scene Assembler provides a user interface to manually specify these occluders in a given environment, in addition to providing a technique for the automatic extraction of the occluder geometry. The user interface is a simple point and click strategy to define a line segment in the 3D environment. This line segment forms an occluder polygon by extruding the line, which is typically close to being horizontal, vertically downwards until it penetrates the terrain. These line segments are automatically derived for a given building by selecting the roof apex.

Once the model has been imported into the Scene Assembler and augmented with semantic information, vegetation, camera paths through areas of interest and a virtual tour guide the file is exported in Collada format. This file is subsequently passed to the EPOCH Scene Renderer for real time navigation and interaction with the natural language processing module.

Within this framework, the Natural Language Processing and Generation module developed by the University of Brighton pro-

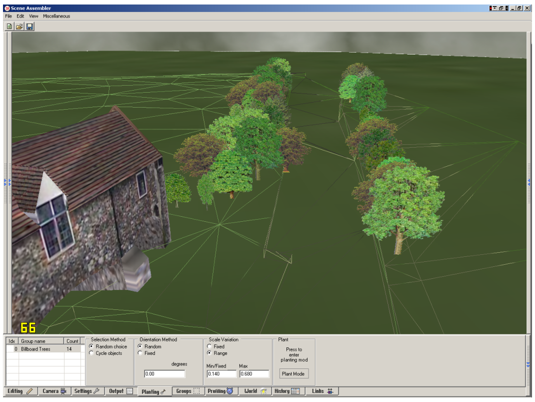

Figure 4: *Scene Assembly Toolkit: Scene Assembler component with a virtual model of Norwich, UK*

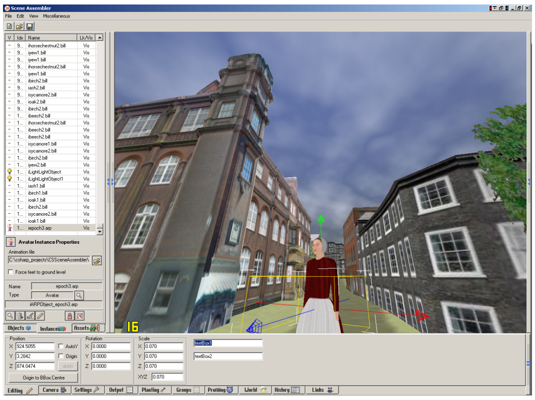

Figure 5: *Inclusion of virtual avatars in a 3D scene of Norwich, UK*

vides natural language interactivity within the environment. This functions as a question/answer system using predominantly natural language. For this, language technology automates the structuring and querying of heterogeneous and semi-structured information in databases which are structured within the CIDOC-CRM ontology [CID08]. The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

In addition, the system uses semantic similarity for producing appropriate answers to user's queries. This similarity is computed using dynamic programming where deletion, insertion and substitution are given a cost. WordNet [Wor08] provides a measure of the cost for substituting one word for another. As a result, the user query Q is compared with a set of n predefined queries. If the most semantically related Q-Qi pair is above a given threshold of confidence, the corresponding answer Ai is selected. Otherwise an Eliza-type of interaction is adopted.

Finally, the potential for integration of the resources and 3D environments produced by both of these frameworks is achieved by using the common open source generic frameworks OpenSG. For this, vhdPLUS includes an OpenSG rendering plug-in and UEA framework is based on the OpenSG rendering framework.

4. Interactive 3D Virtual Environments

In order to illustrate the environments and efficient virtual human simulation produced by both frameworks, two different real-time interactive 3D VR systems environments will be presented in the following sections. Both of these applications recreate a city or town populated by virtual humans for visitors to explore and gain information on the history of the place in an engaging and entertaining way.

4.1. Reviving the ancient city of Pompeii

Its main goal is to simulate in real time a crowd of virtual Romans exhibiting realistic behaviors in a reconstructed district of ancient Pompeii as illustrated in figure 6.

In an offline process, the city is first automatically reconstructed and exported into two different representations: a high-resolution model for rendering purpose, and a low-resolution model labeled with semantic data. Second, the city is populated with crowds of convincing virtual Romans presenting several different behaviors, and thus offering a realistic and varied simulation.

The city reconstruction is based on a CGA Shape grammar [MWH*06] with nine different facade designs derived from archaeological drawings. It contains 4 levels-of-detail, 16 different wall types and three roof styles. Of this grammar 16 variations were automatically generated by combining the facades and roofs with specifically designed color palettes. This number could be arbitrarily increased, but practical aspects of the rendering limited the usable number of materials. The CGA Shape grammar has proven its great flexibility, for instance during the optimization of the levels-of-detail for the rendering process.

There are several buildings in the city model where virtual Romans can enter freely. Some of them are labeled as shops and bakeries, and the characters entering them acquire related accessories, e.g., oil amphoras or bread. These accessories are directly attached to a joint of the virtual character's skeleton, and follow its movements when deformed. We can attach accessories to various joints, depending on their nature. In Pompeii, this variety is illustrated with the amphoras: rich people leave shops with an amphora in their hand, while slaves leave shops carrying them on their heads.

The idea of rich and poor districts is based on age maps that were provided by archaeologists taking part in the EPOCH project. These maps show the age of buildings in the city. Although we do not yet have the building textures to visually express this kind of differences, we have decided to install the rich Roman templates in the most recent districts, while poor people have been established in old buildings. From this, virtual characters know where they belong and while most parts of the city are accessible for everybody, some districts are restricted to a certain class of people: rich Romans in young areas and slaves in poor zones.

As for the crowd, seven human templates have been exploited: a couple of nobles (one male and one female), a couple of plebeians, another couple of patricians and finally, a legionary. These seven templates are instantiated several hundred times to generate large crowds. To ensure a varied and appealing result, per body part color variety techniques are exploited. The resulting application is presented in [MHY*07].

Figure 6: *Crowds of virtual Romans in a street of Ancient Pompeii*

4.2. Touring the town of Wolfenbuttel

This application uses the UEA Scene Assembly Toolkit. It recreates Wolfenbüttel as it once stood during the seventeenth century. This town sits on the Oker river in Lower Saxony-Germany, just a few kilometres south of Braunschweig. Within the 3D interactive application, the user navigates the scene accessing information about important buildings. For this, important landscapes in the scene are annotated with meta data so that visitors can explore related websites or other resources during a virtual tour.

To build the application, 3D modelling packages were used to model the most important buildings of the town. These models and other generic models of houses were imported to the Scene Assembler and the final scene was exported in the COLLADA [Col08] format. The EPOCH Scene Renderer was used for real time rendering of the imported COLLADA file using OpenSG [Ope08b], and provides user interfaces for navigation through the scene.

An interactive female virtual avatar was created using the ARP toolkit in order to have a highly interactive application and to create a more engaging presentation of the history of the town. Thus, the avatar was design to act as a virtual guide which responds to user's questions related to the building and events in the town. This is achieved by modeling appropiate gestures animations and adding natural language understanding to the application.

A "virtual tour" was created by generating interesting route paths in the environment. Six locations have been selected for the user to visit in the virtual reconstruction. The user navigates from one to another by clicking on labels "floating" in the sky. Once at a location, the user can look around, rotating the view or move freely with the use of keys. The user can request more information about any of the six locations in town using the following approaches: i) typing a question or ii) "pointing&clicking" on one of the predefined questions. The user also has access to a webpage when arriving at certain locations.

In this environment seen in figure 7, priority was given to building a mock-up that will allow users with different backgrounds to be part of the full interactive loop (navigation-request-processing-response-navigation) of interaction with a virtual guide. Taking into account contextual (location on the site) information about the user during the interaction provides a first impression of what natural interactive systems can achieve for navigation through Cultural Heritage sites.

Figure 7: *Avatar guide interacting with users of the virtual city of Wolfenbüttel*

The usability and acceptability of the Wolfenbuttel 3D Interactive application was investigated in order to identify the effectiveness and satisfaction of users when interacting with this type of 3D environment. A focus group of 12 users were involved in this evaluation. For this, two different approaches were used: 1) formative evaluation which involves assessing the user interaction by placing representative users in task-based scenarios, and 2) post-hoc usability questionnaire. The requirements, methodology and results of the testing are presented in detail in [REMM*07].

In general testers found the application a good and entertaining representation of a historical place. Avatars were regarded as useful to add life to the environment, although not critical. However, users suggested that the levels of interactivity of the application should be higher. The results show how the multimodality displayed by the system was received with diverse responses, highlighting the struggle in the Virtual Reality domain: that of realism against interactivity. Whilst users wanted realism in the scene (for which they considered avatars to be an important element) they also wanted interactivity above realism. They also preferred those interaction techniques which had been learned from previous experiences with 3D environment, in particular from games.

Users had a more divided opinion on the best way to use these systems for presenting heritage and how they can replace current presentation systems. Testers tend to perceive them more as entertainment. This highlighted the need to make a real connection between the heritage/artefacts in display and the virtual environment. They should enhance the museum/site experience rather than try to be the experience themselves.

To fully evaluate the Scene Assembler's applicability to the cultural heritage domain, a further evaluation was conducted that considered the ease in which the Pompeii scene could be imported into the EPOCH Scene Renderer. To obtain a real time traversal of the Pompeii model in the renderer, the data must first be imported into the Scene Assembler to enable particular areas in the model or buildings to be augmented with metadata. Once the context specific information has been incorporated into the model, the model is saved in Collada format, which is subsequently imported into the EPOCH Scene Renderer. The virtual tour guide is able to process the metadata and via the natural language processing module can interface with the geospatial database and facilitate the real time exploration of the site.

One of the main difficulties in achieving the seamless integration of a city model lies in the lack of standards for a 3D model format, which encompasses all the required information. A 3D model can potentially consist of millions of triangles, with each triangle comprising a set of vertices, texture coordinates and a material. A material may have several associated material properties including diffuse, ambient and specular colours as well as including a filename for a texture map. This type of information may be stored in WaveFront's OBJ format, which is the format that the Scene Assembler imports.

However, this file format lacks the ability to reuse geometry through instancing. Instancing is a particularly useful technique for optimising the rendering of scenes in real time. The technique exploits a fundamental characteristic of a city, that is the fact that there exists many reoccurring structures. However, since the Scene Assembler reads a collection of triangles, this causes every object to be completely described using polygons and the result is a very large memory footprint for the model. The memory footprint is detrimental to the real time performance of the application, but it is also a disadvantage when attempting to export the scene in Collada format. The latter is caused by the fact that the Collada DOM must be held in memory when it is created, leading to prohibitively large memory requirements for generating city environments.

Whilst the Collada format does provide the memory efficient instancing approach, it suffers from a limitation that can effect the visual quality of the scene. It is frequently the case when attempting to convert models between formats that data may be mistranslated or lost. The Collada format does not support a material that consists of texture and material properties. This limitation is apparent when, for instance, attempting to reuse a texture and employing colour modulate with the diffuse colour channel to generate variety. The previously described scenario occurs when texturing a building's roof. A texture may be used to represent the layout of the tiles and the colour values can be visualised by modulating the texture with the colour in the diffuse channel. To overcome this problem many different textures are required leading to an increase in the memory footprint. To circumvent the creation of multiple textures the colour may be encoded into another area of the material attribute in Collada, such as appending the name of the material with the RGB of the colour channel. However, this does not offer a user friendly solution and suggests that more research and development is required before a 3D format is fully identified.

5. Conclusions

The paper has presented results on the Characterize NEWTON within the EPOCH project. Both frameworks presented, vhdPLUS and UEA Scene assembly framework make use of open source generic frameworks to provide vertical middleware for constructing interactive 3D Virtual Environments. Furthermore, the paper has presented example of applications which can be built with this frameworks as a demonstrator of their potential for the Cultural Heritage sector.

The software is available from: [vhd08a] in the case of vhdPLUS and contacting the developers for the Scene Assembly Toolkit.

6. Acknowledgements

This work has been conducted as part of the EPOCH network of excellence (IST-2002-507382) within the IST (Information Society Technologies) section of the Sixth Framework Programme of the European Commission.

Our thanks go to the others involved in the realisation of this work: the research team at the University of Brighton, in particular to Michel Genereux and David Morris. Also to the resarch team at MIRALAB, EPFL and the University of East Anglia. Special Thanks to Phil Flack at UEA for his contribution on the UEA Characterise work. Thanks are also due to the Graphics Research Groups at the Technical Universities of Braunschweig and Graz for their involvement in the development of the Wolfenbüttel application.

References

[CID08] CIDOC CONCEPTUAL REFERENCE MODEL:. http://cidoc.ics.forth.gr/, 2008.

[Col08] COLLADA:. http://www.collada.org/mediawiki /index.php/Main_Page, 2008.

[DWS*97] DUCHAINEAU M. A., WOLINSKY M., SIGETI D. E., MILLER M. C., ALDRICH C., MINEEV-WEINSTEIN M. B.: Roaming terrain: real-time optimally adapting meshes. *IEEE Visualization* (1997), 81–88.

[MHY*07] MAÏM J., HAEGLER S., YERSIN B., MUELLER P., THALMANN D., GOOL L. V.: Populating ancient pompeii with crowds of virtual romans. In *in the 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST'07)* (2007).

[MTP06] MAGNENAT-THALMANN N., PAPAGIANNAKIS G.: Virtual worlds and augmented reality in cultural heritage applications. In *In "Recording, Modeling and Visualization of Cultural Heritage"* (2006), pp. 419–430.

[MWH*06] MUELLER P., WONKA P., HAEGLER S., ULMER A., GOOL L. V.: Procedural modeling of buildings. In *In "Proceedings of ACM SIGGRAPH 2006 / ACM Transactions on Graphics"* (2006), pp. 614 – 623.

[NMTCY07] NADIA MAGNENAT-THALMANN ALESSANDRO FONI G. P., CADI-YAZLI N.: Real time animation and illumination in ancient roman sites. In *In "The International Journal of Virtual Reality"* (2007).

[Ope08a] OPENSCENEGRAPH:. http://www.openscenegraph.org/projects/osg, 2008.

[Ope08b] OPENSG:. http://opensg.vrsource.org, 2008.

[Pon04] PONDER M.: Component-based methodology and development framework for virtual and augmented reality systems. In *PhD Thesis* (2004).

[PPM*03] PONDER M., PAPAGIANNAKIS G., MOLET T., MAGNENAT-THALMANN N., THALMANN D.: Vhd++ development framework: Towards extendible, component based vr/ar simulation engine featuring advanced virtual character technologies. In *In "Proceedings of Computer Graphics International"* (2003), pp. 96–104.

[REMM*07] RODRIGUEZ-ECHAVARRIA K., MORRIS D., MOORE C., ARNOLD D., GLAUERT J., JENNINGS V.: Developing effective interfaces for cultural heritage 3d immersive environments. In *in the 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST'07)* (2007).

[vhd08a] VHDPLUS SOURCEFORGE REPOSITORY:. http://www.sourceforge.net/projects/vhdplus, 2008.

[vhd08b] VHDPLUS SOURCEFORGE WEBSITE:. http://vhdplus.sourceforge.net, 2008.

[Wor08] WORDNET:. http://wordnet.princeton.edu/, 2008.