KINECT VIRTUAL TOURS FOR CULTURAL HERITAGE SITES

Bilo Lwabona, Simba Nyatsanga, Henk van Jaarsveld
Department of Computer Science
University of Cape Town
May 2012
Contents

1. Project Description
2. Problem Statement
   2.1 Research questions
   2.2 Significance of the problem
   2.3 Why is this important?
3. Procedures and Methods
4. Ethical, Professional and Legal Issues
5. Related Work
6. Anticipated Outcomes
7. Project Plan
   7.1 Risks
   7.2 Timeline
   7.3 Required resources
   7.4 Deliverables
   7.5 Milestones
   7.6 Work allocation to team members
1. Project Description

This project is aimed at providing an interactive system for accessing and displaying large 3D models of heritage sites, along with other data such as panorama images and floor plans. The user interaction with the system will primarily be gesture-based, using the Microsoft Kinect sensor, but will also support conventional access using a typical desktop interface. This will enable users to explore these environments in an immersive way and learn about specific characteristics of entities within these heritage sites.

2. Problem Statement

2.1 What is being investigated?

The purpose of this project is to investigate how well virtual environments can accurately mimic real world heritage sites in a way that can be useful for tourists and researchers. One key point is that real-world GIS data is used. To make the proposed system more usable, an investigation will be conducted to determine how natural user interfaces, specifically motion gesture-based interfaces, can be used to interact with the virtual environment. The usability of the aforementioned natural interfaces will be compared with that of a conventional desktop interface, i.e. a mouse and keyboard, to assess the user experience in both methods of interaction.

Users will be able to explore the environments and learn about specific areas of their interest within these environments. Additionally, researchers and other professionals in the field will be able to alter the information. This includes adding, removing and editing contextual information such as images, text, audio and video. Therefore one can classify two types of users, namely general users and specialists.

2.2 Research Questions

Building a 3D model renderer that supports interactive frame rates

The proposed system will be dealing with high-detail, laser-scanned models. We will investigate whether it is possible to build an efficient rendering engine that can display these large 3D models at interactive frame-rates.

Building a terrain renderer that can handle a large number of data points

The terrain data points in question for the proposed project number in the billions. We will investigate whether it is possible to build a renderer that can efficiently handle large data-sets, while accessing the data from a storage medium.

Dealing with uncertainty in data

The laser scans of models may be incomplete (e.g. contain holes). The user needs to get feedback on what type of uncertainty the model contains and where.
Implementing a natural user interface effectively

It will be investigated whether the interface can be implemented in such a way that users are able to explore the virtual environment in a natural and intuitive way. Not only does this investigate the usability of the interface, but it also assesses the appropriateness of gestures for a given interaction command.

Usability of motion gestures compared to desktop interface

The research question being asked here is how a motion gesture-based interface compares to a more commonly used desktop interface. The usability of a user interface is important and thus the efficiency of use and intuitiveness of both interfaces will be compared to each other.

2.3 Why is the problem important?

It is important that more research is done in the field of cultural heritage [19] and virtual environments. It is important to implement a system that can effectively convey information about a given heritage site to a user while keeping the interaction as simple and natural as possible. Additionally, integrating natural user interfaces can make the use of virtual systems more interactive, immersive and fun user experience. Virtual museums do not usually make use of systems that display such accurate data like the one proposed in this document. This makes the project both unique and interesting. All the models and terrain are geo-referenced and the models, recorded using laser scanners, are of extremely high quality. This means that our proposed system will provide both a fun learning tool for casual users, as well as a research tool for professionals.

3. Procedures and Methods

Development Platform

The 3D environments, partially containing vast amounts of vertices, will have to be rendered efficiently. C++ is a fast programming language, which is why it is suitable for this project. Two major graphics rendering APIs are OpenGL and Direct3D (part of DirectX). While OpenGL has the advantage of being compatible with multiple platforms, the Windows dedicated DirectX API has been ahead of OpenGL for several years [10]. The development platform will be Windows, which supports both DirectX and the Kinect SDK, and these have vast amounts of support [11]. The reason we chose DirectX over OpenGL is that DirectX has been various steps ahead in the past years, and it provides control of different types of data, such as video and audio, in addition to its graphics processing capability with Direct3D.

User Interface

The graphical user interface will consist of a dock that displays the corresponding icons for all possible actions at any given instant. The icons will change dynamically, depending on the state that the system is in at the time. This will be useful for both the desktop and the gesture-based interface. The GUI will also indicate the required gestures for specific actions. Furthermore, the
expert version of the system will contain a mode in which guided tours can be created, edited and deleted.

In 3D-computer graphics Arcball [12] is an input technique where users move the mouse in an arbitrary direction to rotate a 3D model around an arbitrary axis. The same mechanism can be used to rotate one’s direction of view while the environment’s orientation remains unchanged relative to the virtual environment’s coordinate system. The W-A-S-D keys on the keyboard are typically used in 3D games for moving around in the environments in 2 dimensions, which are forward, left, back, and right respectively. Occasionally, the Q and E keys are used to move the viewpoint up and down, more commonly encountered in 3D modeling software. This input for translation, in combination with the Arcball for rotating the camera, can be used to navigate through a virtual environment effectively.

**Gesture-based interface**

The Kinect can also be used to navigate through virtual environments. The user can achieve this by moving their hands in arbitrary directions. Combinations of open palms and clenched fists can be used to move the viewpoint and change the camera’s orientation. Alternatively a virtual track ball can be employed to implement the same functionality. Furthermore, both hands can be moved apart in opposite directions to zoom in to the direction the camera is facing. Similarly, the hands can be moved together to zoom out. To move between places far apart from each other, a mini-map could be highlighted, on which the current position is displayed and the desired destination can be selected. Some tasks will be quite similar (e.g. play-back of video and audio, or deleting a document). This will be exploited to minimize the amount of gestures the user has to learn, by employing context sensitive gestures, where appropriate. For example, the same gesture could be used to skip to the next video clip, audio clip, text document, etc. Some generic gestures will also be used. One example of this would be allows the user to exit the current activity, be it reading, watching, listening, etc. , all with the same gesture. At the same time context-sensitive gestures will be considered carefully so as to ensure that sensible motions in one context are still intuitive in another.

**How to render terrain**

The proposed system will use an algorithm called Geometry Clipmaps [3]. Essentially what it does is render part of the terrain that is close to the viewpoint at a high resolution using nested grids around the viewer, and all the other terrain data is compressed and stored on a storage medium. This addresses the question of how to deal with a large data-set as all the other data is only decompressed when it has to be used for rendering. This method also incorporates Level-of-Detail, a common technique in many terrain rendering algorithms to ensure that only the closest part of the terrain, relative to the user, is rendered at high-resolution. This therefore means dealing with a substantially low number of polygons and greatly improving the frame-rate of the renderer. The algorithm uses a simple grid data structure and is therefore easily extendable to a GPU implementation [4], by using the GPU’s rasterizing pipeline to handle the grid of polygons. To add more detail the rendering will utilize a similar clipmap technique called texture Clipmap [5], in which the textures are stored at different levels-of-detail and the appropriate level of detail is chosen depending on how close the viewer is. This technique, combined with exaggerated shading will be utilized to illustrate the detail in the terrain, and for this non-photorealistic
shading model which uses a soft toon-shader [6], will be used.

**How to render 3D models**

The 3D computer models that will be used in the application are provided by the Zamani Project [18]. The models were created using laser-scanners and are of very high quality. Some of the models have a vertex count of up to 7 billion [13], these models could potentially be too complex to render at interactive frame rates.

For each model that needs to be displayed, lower resolution models are also provided. These will be used for Level of Detail (LOD) techniques in rendering. LOD is a powerful technique used to display large data sets by actively swapping the currently viewed high-resolution models with lower quality ones for efficient rendering. Swapping over of the models should be seamless, so as to keep the user's immersive experience uninterrupted.

If there exists some uncertainty in the models, it needs to be visualized. The user needs to be aware that some of the data has been generated and not captured. A simple but affective way of showing uncertainty in certain parts of a model is blurring the texture at the respective areas.

**Prototype**

The desktop interface will be implemented early in the development cycle, so that navigation of the environment is not restricted to the Kinect. This is particularly important for developing components of the system, which are independent of the method of interaction.

A high fidelity prototype will be implemented to evaluate the gesture interface. Fake data for the terrain and 3D models will be used with a partially, or potentially, a fully functional interface to conduct this experiment. As stated above the prototype will be developed in with DirectX and the Kinect SDK. The prototype will be useful for usability testing, which is a key component in iterative improvement. Responsive and intuitiveness are the major aspects of the interface to be evaluated and improved upon. Furthermore, the environment used in the prototype will contain the various types of data (video, audio, text, panorama, floor plans, etc.), such that the visualization thereof can also be evaluated and improved upon. Different visualization techniques will be implemented and compared to assess which is the best for visualizing the information.

The rendering component will be prototyped differently. Immersive elements such as the skymap and other ambient elements such as sound will be less complicated to implement than the terrain itself. Therefore these will be implemented early in the development cycle. The prototype of the renderer will contain a terrain, with an attempted LOD. The efficiency of the LOD will be evaluated with an interactive system. Since the terrain will only have to be re-rendered when the view changes, this can be used to evaluate the efficiency of the renderer.

**Testing Plan**

The high fidelity prototype will be used to evaluate the usability of the system. User experiments will be conducted, and these will be video recorded for analysis. We plan on having fifteen users in the experiment comprising of five experienced Kinect users, five users with some exposure to virtual environments such as 3D video games, and five users with very little, or preferably, no
experience whatsoever in either motion-controls or virtual environments. This sample space will yield valuable information about the system’s usability. These evaluations are going to be used to make improvements on the system’s interface.

The renderer’s efficiency will be tested in several ways to detect any lag. This can be done by changing the camera’s direction of view at different speeds, as the environment is re-rendered when the view changes. This could also be tested by moving around in the environment.

During development it is important to know exactly when the models are swapped with copies of different resolutions. A suitable type of feedback would be a change in colour. To elaborate, a spectrum of colours could be used, where, as an example, lower resolution models are rendered with a certain colour and higher resolution models are rendered with higher saturation. There is an alternative to this approach, should it be unclear which the different stages are. As opposed to saturation, a temperature map (colour spectrum) could be used. High resolution models could be rendered in red (high temperature), where low resolution models would be rendered in blue (low temperature), and the any intermediate resolutions would be assigned corresponding colours of the spectrum in between these two extremes.

We should be able to determine how seamless the transitions are between different levels of detail. Key variables are the amount of different resolutions per model, as well as at which distances the models are swapped. The aim of this specific user testing scheme is to determine how obvious the swapping of models is. These variables will be adjusted during the tests to obtain the best balance of variables.

The basic system should be completed as soon as possible. Furthermore, plenty of iterative improvement should be done, so that the final product is as polished as time permits.

4. Ethical, Professional and Legal Issues

The GIS data, provided by the Zamani project team, is not intended for inappropriate use, such as creating a virtual environment for games, or any other purpose even remotely unrelated to the project. The data could contain sensitive information. Hence, the privacy should not be exposed or violated in any way. Since a lot of user testing will be carried out, the ethics clearance form will also have to be completed.

5. Related Work

Natural user interfaces have been used to explore virtual environments in the past. One such application is Arch-Explore [15], which is a system that makes use of a head mounted display (HMD) to provide an immersive experience when walking through an environment. Another implementation enables users to view 3D architectural models, using the Kinect as an input device [16]. A less related application of the Kinect is one used to browse medical image data [17]. The relevant part here is the types of gestures used to implement a functional system.

Virtual reality has become an important technique for visualization, restoration and preservation of heritage sites. The National Technical University of Athens used AutoCAD
and 3dsMax to create a virtual depiction of the Byzantine Monastery of Dafni in Athens [7]. To ensure accurate depiction, they used accurate orthophotomosaics taken from the monastery to create textures for its virtual counterpart. Many restorations have been done using multi-resolution models in order to have detailed virtual environments, while being space efficient [8]. These include The Etruscan Necropolis of Tarquinia in Italy, Pompeii and Roman forums in Italy as well as Angkor Wat temples and statues in Cambodia. The University of Geneva in collaboration with Yildiz Technical University of Istanbul, constructed a virtual depiction of the Hagia Sophia in Turkey [9]. They used NURBS (Non-uniform Rational B-Spline) modelling to accurately depict the curves in the chambers of the mosque. This modelling technique was also used because of the efficiency it provides for real-time interaction. Global illumination techniques such as photon-tracing were also used for lighting of the Hagia Sophia.

6. Anticipated Outcomes

6.1 System

3D terrain renderer
The geographical surroundings of the heritage sites will be rendered in an efficient, and, ideally, realistic way.

3D model renderer
Depending on the particular heritage sites, the environment may contain man-made structures, such as churches, buildings, etc. These models will be rendered in a way that allows users to move around in the environment and inside the models without experiencing lag. The user will be able to see which parts of the model have been interpolated and which have not.

User Interface
There will be two ways of interacting with the environment. Gestures will be used, detected by the Kinect, to interact with the environment naturally. The alternative interaction method will make use of traditional input devices, namely a desktop interface, composed of a mouse-keyboard setup.

These components, together with a graphic user interface, will form the whole system. The user will make the desired gestures that will then be translated into a transformation in the environment. This is then fed to the renderer to display the desired view. Contextual information like videos, audio, text and panoramas can also be accessed by the user interacting with the system. Additionally, expert users can also add annotations to the site using the desktop interface.

6.2 Measure of Success

Performance: The speed at which the 3D rendering engines will render the terrain and 3D models of the site will be critical in the success of the system. We expect a successful system to
be able to render and swap around the high resolution and low-resolution models alike, at high speeds ensuring good response time.

*Usability:* Not all users are technologically inclined. The gesture-based system is meant to allow for more natural interaction with the virtual heritage site. This therefore means the gesture interface has to be intuitive to use. Alternatively the system will also be usable with a desktop interface.

*Authenticity:* The system will be recreating heritage sites and therefore a measure of success will be how well the virtual sites resemble their real world counterparts. All the data which is going to be used in the application is geo-referenced.

*Accuracy:* The accuracy of the Kinect gestures will also be a measure of success. Users expect a responsive and accurate system that correctly detects their gestures. This is a vital aspect in the usability of the natural user interface.

### 6.3 Expected Impact

A successful implementation of the proposed system could lead to groundbreaking virtual environment technology. The outcome of the proposed project could invoke more use of accurately geo-referenced, laser scanned models in virtual tours. Users of the system will not feel as though they are interacting with a re-created version of the environment but rather a digitalized one. This contributes to the immersive experience a user has when interacting with the virtual environment.

Additionally, the HCI component of the proposed project could stimulate interest in other methods of natural interaction. As an example, the use of touch gesture based systems could be an alternative method of interaction.

### 7. Project Plan

#### 7.1 Risks

<table>
<thead>
<tr>
<th>RISK</th>
<th>SEVERITY</th>
<th>LIKELIHOOD</th>
<th>MITIGATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group Dependency:</strong> The success of the project highly depends on the dedication and competence of every group member. Failure of this due to illness or personal reasons may put the project at high risk of not being completed.</td>
<td>High</td>
<td>Medium</td>
<td>The components of the project are broken down into clearly identifiable sections for each individual to reduce dependency</td>
</tr>
<tr>
<td><strong>Failure or delay in obtaining hardware:</strong> The project is at risk of not being completed if we fail</td>
<td>Low</td>
<td>Low</td>
<td>We own two Kinect sensors. Additionally, the department also has one available for use.</td>
</tr>
</tbody>
</table>
to get the adaptor component that is needed to connect the Kinect to a PC.

<table>
<thead>
<tr>
<th>Risk Area</th>
<th>Probability</th>
<th>Impact</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure to obtain adequate GIS data</td>
<td>High</td>
<td>Low</td>
<td>We are in communication with the clients who will be providing the data for the use of the project. They already have all the data available, and only need to organize it and give us a copy.</td>
</tr>
<tr>
<td>Failure to meet milestones</td>
<td>Medium</td>
<td>Medium</td>
<td>Weekly meetings will be scheduled between the group members to ensure everyone knows what they need to be working on when.</td>
</tr>
<tr>
<td>Complexity of the 3D Renderer</td>
<td>Medium</td>
<td>Low-Medium</td>
<td>A prototype of the renderer will be implemented as soon as possible to ensure enough time for testing and full implementation.</td>
</tr>
<tr>
<td>Conflict between group members</td>
<td>Low</td>
<td>Medium</td>
<td>The group members generally work together. Clear communication channels will be established, as well as weekly meetings to make sure all the progress is clearly communicated to every member.</td>
</tr>
<tr>
<td>Data Loss</td>
<td>High</td>
<td>Low</td>
<td>A strict access control scheme will be enforced among the project members. Multiple copies of the dataset will be stored on non-volatile storage</td>
</tr>
</tbody>
</table>
media. The data providers (Zamani Project) are expected to have multiple backups too.

<table>
<thead>
<tr>
<th>Unclear Scope: Completion of the system highly depends on a clear definition of the scope. Failure to do this may lead to an incomplete system, or a system solving the wrong problem.</th>
<th>High</th>
<th>High</th>
<th>Frequent meetings with the supervisor(s) and possibly with the clients will ensure a clear understanding of the scope of the project.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maintaining data authenticity: The project involves working with authentic dataset from heritage sites. There is risk of false interpretation of the sites by the 3D renderer.</td>
<td>Medium</td>
<td>Low</td>
<td>The implementation of the renderer will ensure that there is no modification of the provided data. Strict access control will also ensure that no member tempers with the data.</td>
</tr>
<tr>
<td>Development Environment The system will be developed using the DirectX11 SDK, which is only compatible with the Windows operating system.</td>
<td>High</td>
<td>Low</td>
<td>Two of the three members own a personal computer that runs on the Windows 7 operating system. The computers in the Honours laboratory also run the Windows 7 operating system.</td>
</tr>
<tr>
<td>Inappropriate use of data by group member(s): Since the data is not meant for inappropriate use, losing it might lead to unwanted legal or ethical issues.</td>
<td>High</td>
<td>Low</td>
<td>Strict access control will be enforced to make sure that no member uses the data in an inappropriate manner.</td>
</tr>
</tbody>
</table>

7.2 Gantt chart
7.3 Required resources

The project will require a Kinect motion sensor for the gesture based interaction with the virtual environment. In order to program using this device, a power/usb adapter is needed. A sufficiently powerful computer is also required to render 3D models in real-time and to allow interaction. The software required for the development of the system is DirectX11 and a C++ compiler. Since DirectX is only compatible with Windows, the computers need to be running this operating system.

7.4 Deliverables
The final system should provide a realistic visualization of heritage sites, displayed in a virtual environment. Models will be accurately placed into this environment using real GIS data as reference. Users can interact with the environment by using either a Kinect sensor or a desktop interface. Interaction with the system will be composed of navigating through the environment, and interacting with area of interests. The system should also provide paths to the different parts of the site using a path finding algorithm. The aforementioned desktop interface should allow expert users to add contextual data to the sites.

7.5 Milestones

<table>
<thead>
<tr>
<th>Milestones</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Proposal</td>
<td>21(^{st}) May</td>
</tr>
<tr>
<td>Project Proposal Presentation</td>
<td>24(^{th}) May</td>
</tr>
<tr>
<td>Initial Project Web Presence</td>
<td>12(^{th}) June</td>
</tr>
<tr>
<td>Initial Feasibility Demonstration</td>
<td>13(^{th}) June</td>
</tr>
<tr>
<td>Background/Theory Chapter</td>
<td>29(^{th}) July</td>
</tr>
<tr>
<td>Design Chapter</td>
<td>29(^{th}) August</td>
</tr>
<tr>
<td>First Implementation and Write-up</td>
<td>19(^{th}) September</td>
</tr>
<tr>
<td>Final Prototype and Write-up</td>
<td>28(^{th}) September</td>
</tr>
<tr>
<td>Final Implementation Completed</td>
<td>3(^{rd}) October</td>
</tr>
<tr>
<td>Outline of Complete Report</td>
<td>10(^{th}) October</td>
</tr>
<tr>
<td>Final Complete Draft of Report</td>
<td>24(^{th}) October</td>
</tr>
<tr>
<td>Project Report Final Handin</td>
<td>31(^{st}) October</td>
</tr>
<tr>
<td>Poster</td>
<td>3(^{rd}) November</td>
</tr>
<tr>
<td>Web Page Completion</td>
<td>7(^{th}) November</td>
</tr>
<tr>
<td>Project Demonstration</td>
<td>8(^{th}) November</td>
</tr>
<tr>
<td>Reflection Paper</td>
<td>11(^{th}) November</td>
</tr>
</tbody>
</table>

7.6 Work allocation to team members
The project will be split into three components, such that each student has a viable project in their own right.

Simba will implement the terrain rendering engine and the panoramas.

Henk will implement the rendering of 3D models of the site and integration of contextual data (videos, text, audio files, floor-plan, etc.).

Bilo will implement the system’s Kinect and desktop interface, as well as the guided tour creator. The outcome of this part of the project can be categorized as a “system development and implementation”

REFERENCES


