Real-time realistic rendering and subtle body movements for cultural heritage virtual actors

George Papagiannakis, Arjan Egges, Alessandro Foni, Nedjma Cadi, Nadia Magnenat-Thalmann
MIRALab, University of Geneva, Geneva, Switzerland

Abstract

This paper proposes two simple and robust algorithmic solutions for two frequently encountered problems in inhabited virtual heritage worlds: a) the 'still-frozen' behavior of virtual actors once their predefined-scripted animation is executed and b) their real-time rendering which is mostly often based on non physically-principled local illumination models. Our described framework thus provides a) a procedural keyframe generator for realistic idle motions blended with the virtual actor's scripted behavior and b) an extension to precomputed radiance transfer for multi-segmented hierarchies for global illumination for real-time. Thus on one hand the visual discrepancy between the actor's dynamic illumination model and the environment static high dynamic range lightmaps is greatly minimized; on the other hand, once the virtual actors complete their speech, they continue realistic subtle body motions in a non-repetitive manner, waiting for their next verse instead of stopping in a specific pose. Our case study on verses of ancient Greek tragedies in the virtually restituted ancient theatre of Aspendos illustrates the above integrated algorithms and overall methodology.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

1. Introduction

Virtual worlds [MK94] and their concept of cyber-real space interplay invoke such interactive digital narratives that promote new patterns of understanding. However, the "narrative" and ‘interactive’ parts, which refer to a set of events happening during a certain period of time, and providing aesthetic, dramaturgical and emotional elements, objects and attitudes [NM00], are still an early topic of research. By mixing such aesthetic ambiences with virtual character augmentations, and by adding a dramatic tension to the presented content, we have very recently witnessed these narrative patterns to develop into an exciting new edutainment medium. This paper proposes two simple and robust algorithmic solutions for two frequently encountered problems in inhabited virtual heritage worlds: a) the 'still-frozen' behavior of virtual actors once their predefined-scripted animation is executed and b) their real-time rendering which is mostly often based on non physically-principled local illumination models. Our described framework thus provides a) a procedural keyframe generator for realistic idle motions blended with the virtual actor's scripted behavior and b) an extension to precomputed radiance transfer for multi-segmented hierarchies for global illumination for real-time.

1.1 Overview

Our aim is two-fold: to minimized on one hand the visual discrepancy between the actor's dynamic illumination model and the environment static high dynamic range lightmaps; on the other hand, once the virtual actors complete their speech, to continue realistic subtle and salient body movements in a non-repetitive manner, waiting for their next verse instead of stopping in a specific pose. Our case study on verses of ancient Greek tragedies in the virtually restituted ancient theatre of Aspendos illustrates the above integrated algorithms and overall methodology. The structure of this paper commences by briefly reviewing the most relevant works in this area of virtual heritage, in section 1.2. In Section 2 we present the modeling processes that have been deployed to assist in the creation of the main elements constituting such interactive virtual worlds. In Section 3, we present the characteristics featured by our framework, which provides the necessary central functionalities that are indispensable to allow the handling of the exponential complexity inherent to the simulation of virtual character dramas that traditional rendering-centric VR systems cannot anymore handle. In Section 4 we describe our new MR illumination model for virtual animated characters. Finally, in section 5 we present the technologies and tech-
techniques that have been developed to allow the simulation of advanced real-time virtual humans exhibiting personality and emotion simulations. Our case study of the inhabited VR site of ancient Aspendos, together with the discussion and conclusions, are discussed in section 6.

1.2 Previous work

Virtual actors have already been synthesized in real time virtual heritage applications [DB00], [BFK*01], [PSO*05], [RD05]. However in most of these works the 'stillness' problem is apparent when an actor finishes his/her keyframe or masked by continuously repeated keyframes. [EPM06] proposed an algorithm for idle motions with posture shifts that our work applies in a theatrical setup of an excerpt from an ancient Greek tragedy where actors where not allowed exaggerated variation between verses but also not absolute stillness either. We extend this algorithm by allowing subtle idle body movements and by integrating it in a python scripting control environment, firstly described in [PSO*05]. This enables digital artists to direct the story-telling experience by controlling the timing and execution of predefined verses, blended in real-time with realistic idle motions.

Modeling of currently non-existent complex virtual heritage sites and monuments is still an open issue, as the modeling process cannot benefit from image based modeling and other computer vision and photogrammetry based techniques [PV02]. In this work we provide our modeling approach for realistic, physically-principled complex ancient edifice, tailored for real-time visualization.

High fidelity rendering of virtual humans has recently been enhanced with the adoption of Precomputed Radiance Transfer methods [SKS02], [KLS05]. However none of these works consider multi-material, multi-segmented, skeleton-based skin deformations of virtual actors, or general dynamic scenes with numerous degrees of freedom and dynamic lights. We integrate the observations from [EPM06] allowing for diffuse global illumination for real-time and correct self-shadowing effects for the HANIM [H06] virtual human standard.

2. Modeling virtual worlds for real time (RT) interactive applications

The deployment of virtual reconstruction methodologies and techniques, aimed at the achievement of photorealistic results for an interactive 3D real time experience in a VR inhabited worlds, have been applied to the virtual restitution of the historical site of Aspendos. The necessary methodological and technical solutions required to realize the visualisation of the Roman theatre of Aspendos, along with the inclusion of 3D fully animated Real Time virtual humans re-enacting an ancient Roman play, have been implemented in a real-time VR-AR Framework. Furthermore, specific modelling, illumination and real time rendering strategies, along with the design choices that were operated regarding both the preparation of the textured 3D scene with the animated virtual actors, and concerning the optimizations aimed at the creation of a viable model suited for the needs of a real time interactive visualization, have been carefully considered. For the realization of the 3D model of the Aspendos site, polygonal modelling techniques were generally preferred. Critical parts of the model, such as the Cavea, due to its extremely complex geometrical features, were given special attention to keep an acceptable trade-off between visual accuracy and the performance of the final real time simulation. In order to model the Aspendos site virtual environment, accurate topographic data was provided in the form of a 2D elevation map of the area featuring elevation lines every 1 meter. After the repositioning of these splines in a 3D space, it was possible to build a high polygon 3D mesh (80.000 polygons) from which a greyscale elevation map has been extracted to procedurally generate a diffuse textures using distribution parameters such as relative position of the polygons, their orientation and slope (Figure 1).

High fidelity rendering of virtual humans has recently been enhanced with the adoption of Precomputed Radiance Transfer methods [SKS02], [KLS05]. However none of these works consider multi-material, multi-segmented, skeleton-based skin deformations of virtual actors, or general dynamic scenes with numerous degrees of freedom and dynamic lights. We integrate the observations from [EPM06] allowing for diffuse global illumination for real-time and correct self-shadowing effects for the HANIM [H06] virtual human standard.

To further increase the visual impact of the restituted environment, the addition of trees and vegetation has been included with the implementation of billboard techniques (as depicted in Figure 2). Such billboards, representing a total of more than 30 different types of trees, have been prepared using as starting point real images of trees and bushes that have been edited to isolate their shape from their original background, and to add an alpha channel to allow their visualization with transparency effects.

In order to create the texture tiles to be mapped on the 3D geometry of the restituted theatre, the creation of a library of the main identified diffuse material textures present on the site has been made using as a base the on-site taken high resolution digital photographs: 150 different color corrected textures that can be tiled seamlessly on the objects of the 3D scene have therefore been prepared and assigned to the model's surfaces.
Figure 3: Samples from the texture gallery (left), 3D scene building with and without textures (center and right)

To further improve the overall final visual impact of the 3D interactive real-time simulation of the Aspendos site, the addition of diffuse and hard shadows cast by the sun plays a central role. At an earlier stage of the virtual restitution a full texture baking approach of a pre-computed radiosity solution has been used (Figure 3). This approach however implies the creation of one texture per object: the bigger is the object, the higher the resolution of the generated texture has to be in order to avoid visual discontinuities. Therefore, the risk to overload the video memory, either with excessively high resolution textures or by too many generated textures, is addressed. To overcome these restrictions the adoption of a light-map and diffuse-map real time multi-texturing approach has been adopted due to the fact that the diffuse textures can be tiled, and therefore their resolution becomes independent from the size of the object upon which they are mapped. Secondly the light maps can be downsized to very low resolutions without compromising the overall visual impact of the simulation, thus eliminating the video memory overload issue. In order to visually simulate a more convincing virtual illumination of the scene, the use of High Dynamic Range Image Based Lighting has also been implemented to create such lightmaps, as shown in Figure 4.

Figure 4: 3D final textured model of the theatre of Aspendos Illuminated with the final light probe

However, even though a first illuminated model of the site featuring HDR IBL has been produced during earlier tests using as source for the illumination photographic light probes, the Aspendos virtual model had to be simulated under specific lighting conditions and at specific dates and hours of the day. Thus, the use of virtually generated light probes, allowing an arbitrary positioning of the direct sunlight, by means of parameters such as location, date and time, has been implemented. In order to produce suitable virtual light probes, two approaches were used: the combination of several LDR panoramic images rendered at different exposures, and the direct computation of a synthetic HDR image. Having achieved acceptable results with either method, a virtual light probe featuring the lighting conditions at the site of Aspendos at the 9th of June 2004, 15.00 pm, under an unclouded sky, has been produced and used to compute a global illumination solution for the texture baking process and the creation of the lightmaps to be used for the real time simulation.

The final computed illumination solution has then successfully been used to generate all the lightmaps that subsequently have been assigned to the self-illumination channel of their associated surfaces in the 3D scene in order to modulate the luminosity of the tiled material textures for the real time simulation. Thus, the previous 3D model based on a full texture baking approach of the lighting simulation has been replaced by the new model employing a lightmap based approach. Such approach allows, in fact, the dissociation of the lighting information from the material textures, and consequently allows for an overall better visual quality of the real time rendered surfaces and a reduction of the total weight of all the textures loaded into memory to perform the simulation.

In order to optimize the performance of the real time application, all the lightmaps have been also down-sampled to a size of 256x256 pixels and have been blurred to remove possible compression artifacts due to their small size in comparison to the spatial extension of their assigned surfaces: the final visual quality achieved, in the form of screenshots taken while running the real time interactive simulation, is illustrated in Figure 5.

Figure 5: Real time simulation without lighting information (left), Previous 3D real time full texture baked model (center), screenshots from the final 3D real time simulation of the new model using lightmaps (right)

3. Platform choice for inhabited virtual worlds

3.1 MR (Mixed Reality) character simulation framework design

Our MR system is based on the VHD++ [PPM*03], component-based framework engine, which allows quick prototyping of VR-AR applications featuring integrated real-time virtual character simulation technologies (a schematic overview of the system is depicted in Figure 6). Its key innovation is focused in the area of component-based framework, which allows the plug-and-play of different heterogeneous human simulation technologies such as: Real-time character rendering in AR, supporting real-virtual occlusions, real-time camera tracking, facial simulation and speech, body animation with skinning, 3D sound, cloth simulation and behavioral scripting of actions.

The main design principle was to maximize the flexibility of the system while keeping an efficient real-time performance. The different components constituting the framework may be grouped into the two following main categories:

- System kernel components, which are responsible for the interactive real-time simulation initialization and execution.
- Interaction components, which are driving the external VR devices and are equally providing interactive
scenario authoring, triggering and control through various implemented GUIs.

Finally, the content to be handled and used by the system to create and drive the virtual world built on top of it has to be specified. Such content may be classified into the two following main building block categories: a) Static content and b) Dynamic content, such as models of the 3D scenes, virtual humans, objects, animations, behaviors, speech, sounds, python scripts, etc.

3.2. MR Framework operation for character simulation

The software architecture is composed of multiple clearly defined software components called services, which cover precise responsibilities and provide specific functionalities. Such services take care, for instance, of the rendering of the simulated 3D scenes and the playing of any associated sound or music, they process the inputs from the external VR devices and they control and drive the animation of the 3D models, including complex virtual character skeletal animation, skin deformation and cloth simulation. Additionally, they are also responsible to maintain the consistency of the simulation and to control the interactive components of the scenarios, which can be modified at run-time through the real time parsing of python scripts. Such a mechanism allows the user either to freely modify and interact on the fly with the simulation, or to load pre-made scripts that describe specific state changes to be applied to the simulated elements. In order to keep a good performance while performing such complex and diversified operations, the system utilizes four threads. One thread is used to manage all the updates concerning the services which need to be computed, such as human animation, cloth simulation or voice and sound management. A second thread is used for the 3D renderer, which receives from the current scene-graph all the information concerning the objects that must be rendered and the associated shaders. The third thread features the capability to blend keyframe and procedural animations in real time. Finally, the last thread is constituted by the python interpreter itself. Such thread allows the creation and triggering of Python scripts in order to permit the direct manipulation of the application at the system level, such as, for instance, allowing the generation, loading and application of scripted behaviors, and specific actions, to the virtual humans included in the simulation (Figure 11).

Through the implementation of such approaches and design choices, the MR system presented in Figure 6 succeeds in featuring an immersive real-time interactive simulation supplied with proper information in course of the simulation. That is why content components are much diversified and thus their development is an extremely laborious process that inherently involves long and complex data processing pipelines, multiple recording technologies, various design tools and custom made software. The various 3D models included in the virtual environments, such as for instance virtual humans and auxiliary objects, have to be created manually by 3D designers. The creation of complex 3D virtual humans requires the recording of motion captured data to permit a realistic skeletal animation as well as the constitution of a database of real gestures for facial animations. Sound environments, including voice acting, needs to be recorded in advance based on the storyboard describing the scenarios to be integrated in the simulated virtual world. For each particular scenario, a dedicated system configuration data specifying the operational parameters of the system is defined. Such configuration encompasses both the parameters required to drive the simulation of the physical environment and the parameters used to control the external VR devices. Moreover, the preparation of the scripts defining the atomic behaviors of the simulated elements, such as for the virtual humans, is equally carried out. These scripts can modify any data in use by the current simulation in real-time, thus allowing the simulation to run whilst modifications are performed on the fly.

4. Real-time illumination models for interactive virtual inhabitants

The main contribution of this effort is to propose a new MR [PSO*05] physically correct illumination model based on Precomputed Radiance Transfer [SKS02], [KLS05] extended allow the handling of complex animated deformable virtual humans, and employing ‘real light’ information
captured as light probes [PFT05] to light in real time the simulated 3D elements. The application of such approach allows the implementation of a believable MR illumination registration between the real elements of the simulation and the virtual augmentations. It equally permits to achieve an interactive exposure matching between the real light, as perceived by the real AR camera, and the virtual exposure of the acquired, simulated Area Light which illuminates the augmented scene. The following figure illustrates and example of such approach and is completed with a comparison to previous attempts.

Figure 7: Our extended precomputed radiance transfer for multi-segmented virtual actors illumination model (top) as opposed to standard Phong (bottom)

In order to enhance the consistency between the illumination of the real scene and the virtual characters, we have chosen to extend and adapt the Precomputed Radiance Transfer (PRT) illumination model (Figure 7), to allow its application to the multi-segmented, deformable and animatable hierarchies that our animation system (MIRAnim) is controlling. The main issue that we have encountered and solved, was to allow the multiple segments constituting the virtual skeleton of the 3D virtual humans, due to the presence of different joints, segments, cloth, hair meshes, to respond to a high dynamic range area light captured from the real scene with a method similar to the one described in [EPM06].

Since the final shading information is stored per vertex in the H-Anim virtual human hierarchy, it is independent from the underlying animation approach. Therefore, our virtual human animation system can easily be integrated within our MR framework to drive the animation of the skeleton of the simulated 3D deformable virtual humans. Moreover, it equally allows the inclusion of correct shading information and effect employing the blending schedule defined in Section 5.

Finally, in order to validate the results of our illumination registration in VR based on the extensions from [EPM06], in this work we have compared our real-time rendered virtual characters with offline global illumination simulations of the same meshes. Various offline global illumination techniques were utilized such as Radiosity and Ambient Occlusion with area lights in order to quantify the visual differences phenomenologically. Figure 8 illustrates our comparisons.

Figure 8: Our extended precomputed radiance transfer for multi-segmented virtual actors illumination model (left) as opposed to offline Radiosity based (middle) and sky-light ray-tracing (right)

The offline rendering results have been computed utilizing the 3DSMax rendering engine and the same HDR light probe light was employed in all tests, as created by [PFT05]. Most evident differences occur on the borders of different materials and segments such as the himation on top of the chiton; e.g. in the offline renderings mesh subdivision allowed for less evident self-shadowing artifacts. However, our real-time approach allows for different dynamic lights to be employed in real-time as part of diffuse real-time precomputed radiance transfer.

5. Simulating realistic virtual humans for real-time applications

5.1 Virtual Human Animation

There exist many techniques for animating virtual characters. Two very commonly used techniques are:

**Key-framing**: starting from a set of key-frames, either manually designed by an animator or automatically generated, an animation is constructed by using interpolation techniques. Although this method results in very flexible animations, the realism of the animations is generally low, unless a lot of time and effort is invested.

**Pre-recorded animations**: an animation is recorded using a motion capture/tracking system such as Vicon or MotionStar. The animation realism is high, but the resulting animation is usually not very flexible.

A method like the Principal Component Analysis (PCA) can determine the occurrence of dependencies between variables in a data set. The result of the PCA is a matrix constructed of a set of eigenvectors, which converts a set of partially dependent variables into a new set of variables that have a maximum independency. The PC variables are ordered corresponding to their occurrence in the dataset. Low PC indices indicate a high occurrence in the dataset, while higher PC indices indicate a lower occurrence in the dataset. As such, PCA is also used to reduce the dimension of a given set of variables by removing the higher PC indi-
ces from the variable set. We will use the results of the PCA later on to synthesize the dependent joint motions. For our analysis we perform the PCA on a subset of H-Anim joints. In order to do that, we need to convert each frame of the animation sequences in the data set into an N-dimensional vector. In order to represent the rotation transforms, we use an exponential map representation [EMM04, Gra98]. In this type of representation, a rotation can be represented by a 3-dimensional vector \( r \), and as a rotation with angle \( |r| \) around axis \( r \). The exponential map representation of a rotation is very useful to achieve fast motion interpolation [PR97], due to the fact that it allows the performing of linear operations on rotations. In our case, the linearity of the exponential map representation is crucial, since the PCA only works in the linear domain. In fact, any rotation matrix can be written in the exponential map representation, and any exponential map representation (modulo 2 \( \pi \)) is a rotation. Grassia [Gra98] provides an extensive overview of the advantages and disadvantages of various representations of rotations, including the exponential map. In our approach, one posture/key-frame is represented by 25 joint rotations and one root joint translation, resulting in a vector of dimension 78. We have then applied a PCA on a large set of motion captured postures, resulting in a PC space of equal dimension.

6. Case studies

6.1 Simulating interactive virtual plays in ancient theatres

The ERATO INCO-MED EU Project (http://www.at.oersted.dtu.dk/~erato/) aims at re-enacting ancient theatrical plays in virtually reconstructed ancient Theatres. Specifically, we were involved in the reconstruction of the ancient Hellenistic theatre of Aspendos in Minor Asia, Turkey and the simulation of historically coherent 3D virtual actors, re-enacting parts of ancient Greek dramas. The basic scenegraph OpenGL renderer OpenSceneGraph implemented in our VHD++ framework provides both advanced particle system simulation and occlusion and view frustum culling that have been both successfully integrated and tested in the ERATO platform. The aggressive view frustum culling allows the final scene of 300,000 polygons with 1GB of total textures size to be rendered at ~20fps, according to the view frustum and part of the visible scene within the virtual camera frustum. Such scene includes the full 3D model of the theatre, featuring radiosity processed lightmaps applied at material level to simulate global illumination trough a multi-texturing approach, and the 3D real-time virtual actors illuminated through PRT (precomputed radiance transfer).

Regarding the reconstructed theatre, offline light-mapping rendering methods have been utilized in order to allow the viewing of the theatre under the static area light created using the Aspendos virtual Light probe. The main reason for the implementation of a static lightmap solution was the fact that the reconstructed model exhibited a very high polygonal count and a high number of required Texture memory, thus making any real-time algorithm not destined for such heavy scenes, difficult to apply. Figure 10 illustrates a final example featuring 3D Virtual Humans rendered via our extended precomputed radiance transfer algorithm for illumination registration [EPM06] in the virtual inhabited world of the Aspendos site, employing the same area light used to to statically illuminate the mesh of the theatre of Aspendos. Finally the virtual actors and their clothes were modeled based on the work described in [MCS*04]. The verses are spoken in ancient Greek and the excerpt has been chosen from the tragedy “Antigone” of Sophocles. Figure 11 illustrates an example of the integrated script allowing for motion captured keyframe animations, blended with subtle, idle body movements together with speech simulation.

6.2 Conclusions and Future Work

With the extensions and algorithms that we propose for virtual actor salient body language, physically-principled radiance transfer and simulation, coupled with a complete real-time framework for character simulation, we aim to provide more believable dramaturgical notions for heritage inhabited virtual worlds. Such notions could extend further research in virtual heritage and develop it as an exciting edutainment medium.

In our next steps would like to provide a unified Precomputed Radiance Transfer based solution for both large static, multi-material, multi-mesh environments, as well as virtual actors. Currently in our approach only virtual actors
benefit from PRT due to the complexity of manually specifying Occluders, Receivers [EPM06] in large, multi-material hierarchies. In the particular case of diffuse pre-computed radiance transfer function estimation, the work from [PLK05] could contribute in faster evaluation and rotation of spherical harmonics basis function coefficients. Finally, the recent work from [MWH*06] could provide a rule based solution for procedural automatic modeling of large, non existent heritage edifices.

7. Acknowledgements

The work presented has been supported by the EU IST EPOCH IST-2002-507382 project and the EU INCOMED ERATO ICFC-502A3PR03 project.

References


Figure 10: Screenshots from the final 3D interactive real time simulation featuring our adapted Precomputed Radiance Transfer illumination model for deformable characters and a diffuse map-lightmap approach for static surfaces

```python
animationService.start_player()
animationService.activateAction_body(Antigone,"antigone_idle")
# Creon Monologue (Antigone is on subtle-idle movements during Creon’s monologue)
animationService.activateAction_body(Creon,"creon_wrk")
voiceService.activateAction(Creon,'CreonSpeech.Cam01CreonP1',1.0)
animationService.waitUntilActionHasFinished_body(Creon,"creon_wrk",-4.0)
animationService.activateAction_body(Creon,"creon_idle")
#Antigone answers (Creon is on subtle-idle movements during Antigone’s speech)
animationService.cutOffAction_body(Antigone,"antigone_idle",3.0)
animationService.activateAction_body(Antigone,"antigone_wrk")
voiceService.activateAction(Antigone,'AntigoneSpeech.Cam01AntigoneP1',1.0)
animationService.waitUntilActionHasFinished_body(Antigone,"antigone_wrk",-4.0)
animationService.activateAction_body(Antigone,"antigone_idle")
```

Figure 11: Python script allowing for two actors (Antigone, Creon) and their predefined speech (voiceService), body animation (animationService) as well as subtle body movements (idle)