Journal of
Virtual Worlds Research
jvwresearch.org ISSN: 1941-8477

Volume 2, Number 3

Technology, Economy, and Standards.

Community
Creation
Commerce

Artwork by Anshe Chung Studios

Volume 2, Number 3
Technology, Economy, and Standards
October 2009

Sponsored in part by:

TEXAS DIGITAL LIBRARY
department of RADIOTELEVISIONFILM
UNIVERSITY OF TEXAS AT AUSTIN
siRc THE SINGAPORE iNTERNET RESEARCH CENTRE

The Journal of Virtual Worlds Research is owned and published by:

virtual worlds research consortium

Journal of Virtual Worlds Research
jvwresearch.org ISSN: 1941-8477

Volume 2, Number 3
Technology, Economy, and Standards
October 2009

The Role of Interoperability in Virtual Worlds, *Analysis of the Specific Cases of Avatars*

By Blagica Jovanova, Marius Preda, Françoise Preteux
Institut TELECOM / TELECOM SudParis, France

Abstract

In this paper we present current trends in several activities related to avatars. We provide a detailed survey of research literature for avatar appearance modeling, deformation control, and animation. We also introduce several standards, recommendations, and markup languages treating different aspects of avatars from visual representation to communication capabilities. Finally, we shortly introduce the current developments of MPEG-V related to avatars, a recent MPEG standard aiming to provide an interchange format for virtual worlds.

Keywords: 3D graphics; virtual characters; modeling and animation; interoperability; MPEG-4.

The Role of Interoperability in Virtual Worlds, *Analysis of the Specific Cases of Avatars*

By Blagica Jovanova, Marius Preda, Françoise Preteux
Institut TELECOM / TELECOM SudParis, France

Human-like representation is a practice that goes deeply in humanity's history. For all the visual arts, including painting and sculpture, representing humans reveals the interest. Highly realistic or metaphoric representations of humans are now famous pieces of art. In the modern era, with the evolution of new technologies, the physical support for such representations smoothly changes. The first step in this evolution was first achieved by cinema, which introduced the concept of time. Similar as in the traditional theatre, for cinema the representations are not static anymore, but rather evolve in time and move on the screen. The first 2D cartoons were then produced, bringing up a new and revolutionary concept: animation. While in a static representation depicting life requires artistic skills, when performing animation, by only adding motion it is possible to simulate lively objects or environments. The problems are here related to the need to provide static representation at very dense time samples. The evolution of computer science solves some of these problems, turning digital synthetic content into a new form of art.

Within the large family of digital synthetic content types, avatars have a specific place. The reasons relate strongly to psychological and sociologic motivations (Georges, 2009) of humans to dispose of a visual representation in order to position with respect to the others and to the environment. While this property was extensively exploited in the past in successful games, the avatars representating the player in the virtual world and at the same time the interface with it, the current trend is more and more present in the recent developments of Internet, largely known by the name Web 2.0. Here the user becomes part of the big picture— he/she is an active participant that can modify, append, comment, and in general interact with the content. In many cases the user's presence in the digital world is not required to be visually signaled; only the effects of his or her intervention is observable (i.e. adding some comments on a web-page or editing a section of a Wikipedia document). In some other cases, with real-time requirements such as collaborative work or presence in 3D virtual worlds, the visual representation (i.e. the avatar) is a communication vector in itself—it helps in identification, differentiation, and identity protection. In addition, the avatar is a facilitator of communication, informing the others about status and availability.

It is more and more evident that Internet will evolve from a repository of information to a dynamic and lively place where people communicate with each other and jointly interact with the content, and where time, presence, and events become important. It will contain more and more functionalities, copying, extending and enriching the ones from the real world and inventing new ones. In this context, having a visual representation of users in the form of avatars containing personal information, history, personality, skills, etc. becomes necessary. However, today's Internet (in the sense of its initial definition as interconnected sites) is not yet this lively place. There is no interoperability yet between different websites implementing Web 2.0 features—besides a very thin layer of re-using IDs (e.g., OpenID) or aggregating different sources.

On the other hand, 3D Virtual Worlds (3DVWs) became a reality over the last few years. Initially conceived for social purposes being a support for communication (i.e. chatting) and offering awareness on the presence and sometimes the mood of the interlocutors, 3DVWs are now reaching a milestone: the technology for creating, representing and visualizing 3D content becomes available and widely accessible. One facilitator is the fast development of high performing 3D graphics cards (by the likes of Nvdia, ATI) and their availability in ordinary computers—a trend driven by the powerful market of computer games—making almost any Internet user a potential player of 3DVW.

After an enormous step towards the awareness and democratization of virtual worlds provided mainly by the marketing success of Second Life, VWs are now looking for sustainable business models. The most probable situation is that, in the near future, several virtual worlds will be available, offering complementary functionalities and user experiences. The issue of interoperability between them or at least re-usability of assets, avatars, and media content will become more and more important. Standards for representation of graphics and media assets are now available, MPEG being a very active community in providing tools for compressing them. MPEG-4, one of the richest MPEG standards in terms of functionalities, contains specifications of bitstream syntax for audio, video, 2D, and 3D graphics objects and scenes. It can be used as a base layer for ensuring interoperability at the level of data representation. However, only representing the media is not enough for ensuring interoperability in VW. Recently, the MPEG committee identified addition needs and started a new work item, called MPEG-V-Information Exchange for Virtual Worlds with the goal of standardizing metadata that, combined with media representation, will ensure a complete interoperable framework. Part four of this standard deals exclusively with avatars and virtual goods.

In the last two decades, the research community actively worked on developing tools for creation, representation, animation, transition, and display of avatars and several standards and recommendations were created to represent graphics objects and, in particular, avatars. In Section Two we introduce a review of avatar-related research. Let us note that current VW are populated with other types of objects than avatars, which are not directly driven by end-users. The modeling and animation of some of them (especially animals) can be very similar to the techniques used for human avatars. However, behind this type of objects there is no end-user, only computer programs to drive them. From the perspective of the current paper, an avatar can be whatever 3D object that represents and is driven by the end-user. In Section Three we introduce several existing standards and markup languages and show why they cannot provide a complete solution for ensuring interoperability at the content assets level between VWs. While some standards such as MPEG-4 or H-Anim offer a complete set of tools for representing geometry, appearance, and animation, they fail to attach semantics on top of them. On the other hand, several markup languages such as VHML (Virtual Human Markup Language) or HumanML only describe some high-level features (e.g. emotions, interactions) and do not offer a format for interchanging avatars between applications. Based on an analysis of existent VWs and popular games, tools, and techniques from content authoring packages, together with the study of different virtual human related markup languages, we derived a full XML description of avatars, currently retained in the standardization process of MPEG-V and introduced in Section Four. The main elements of the schema and its capability in representing content compliance with the current VWs are presented. In the last Section we conclude our contribution and provide directions for future research.

State of the Art in Avatars Technologies

As for the general case of graphics assets, the most problematic and expensive operation when considering avatars is their creation. The day to day experience in observing human beings makes the human brain a powerful system able to observe without effort any non-natural effect of avatar modeling or animation. With respect to the avatar appearance there are two main approaches: realistic modeling and cartoon-like. While the first is judged with respect to its capability to copy the reality, for the second the interest is in deforming it (i.e. caricatures, strong appearance effects with the goal of emphasizing psychological characteristics). However, in both cases, the animation of the avatars must be as natural as possible. In the last twenty years, several models were proposed for animating avatars and the most promising are the ones trying to simulate the bio-mechanical structure of the real body, based on skeleton and muscle layers. In the remaining part of this section we introduce the major trends in avatar research literature with respect to modeling and animation.

Avatar Modeling

Two different courses can be chosen in order to build a virtual character according to the researched appearance of the virtual character (cartoon-like or realistic), and depending on the technology available to the designer.

On the one hand, the designer can build interactively the model's anatomical segments and set up the model hierarchy. In addition to creating the geometry and texturing it (both representing the avatar appearance) it is also needed to set up the skeleton and link it to the mesh. Several authoring tools and geometry generating mechanisms make it possible to model a virtual character (3DSMax, Maya). The main drawback of this method is that the result is strongly dependent on the designer's artistic skills and experience. In addition, this procedure is tedious and time-consuming.

On the other hand, a faster and proven method is the use of 3D scanners. Contrary to computer-aided design, the aim of 3D scanning is to create an electronic representation of an existing object, capturing its shape, color, reflection or other visual properties. In its principle, 3D scanning is similar to a number of other important technologies (like photocopying and video capture) that quickly, accurately, and cheaply record useful aspects of physical reality. The scanning process can be structured according to the following steps: acquisition, alignment, fusion, decimation and texturing. The first step aims at capturing the geometric data of the 3D object by using a dedicated scanning device. Depending on the type of scanner used, the execution of this phase can vary considerably. Either a single scanning is enough to capture the whole object, or series of partial scans (called range maps) are needed, each of them covering a part of the object. In the latter case, range maps taken from different viewpoints have to be aligned, which is the task of the second step. This procedure can be completely automatic if the exact position of the scanner during each acquisition is known. Otherwise, a manual operation is needed to input the initial placement, and then the alignment is performed automatically. Once aligned, the partial scans need to be merged into a single 3D model ("fusion" step). Because 3D scanners provide a huge amount of data, a "decimation" step is required. For an effective use of the model, one has to reduce the size of the acquired geometric information, especially of the less significant parts of the object. Decimation software can be based on edge collapsing [Ronfard96] and error-driven simplification [Schroeder92]. The last step, "texturing", is not mandatory for applications such as simulations in a virtual environment but, for a large array of objects, additional information about the real appearance of the object must be provided. This is usually achieved by texturing the final model, using pictures taken during acquisition. The pictures are first aligned to the geometry (manually or automatically) and then mapped to the model.

A recent trend supported by the development of vision systems consists of capturing real persons by using one or several cameras and reconstructing or modifying an existing template by using real measurements. In the case of monocular images such as in Hilton (1999) the geometry obtained is mapped on a previously created model, providing a cheap and useful approach for automatic modeling. By using stereo or general multi-view systems, the 3D geometry may be recovered more accurately. One method consists of computing the disparity map from a stereo pair of images and some local differential properties of the corresponding 3D surface such as orientation or curvatures. The usual approach is to build a 3D reconstruction of the surface(s) from which all shape properties will then be derived. In Devernay (1994) a method directly using the captured images to compute the shape properties is proposed. Some more recent techniques, such as Nebel (2002) can obtain both the geometry and the skeleton. When more cameras are used, such as described in D'Apuzzo (1999), 3D least squares matching techniques can be employed to obtain the geometry and skeleton.

Introducing anthropometry (studying and collecting human variability in faces and bodies) in computer graphics (Dooley, 1982) made possible the creation of a parametric model defined as a linear combination of templates. The basis is extracted from large databases including human measurements such as NASA Man-Systems Integration Standard [NASA95] and the Anthropometry Source Book [NASA78], and several methods in exploiting it are provided (Seo, 2002; DeCarlo, 1994]. An alternative method consists of defining a default model a priori and declaring on it anthropometric parameters with which the model can be deformed; the deformation on the model can be rigid, represented by the corresponding joint parameters, and elastic, which is essentially vertex displacements. Then a dataset with relation between the value of these parameters and shapes of corresponding models is created; from this dataset interpolators are formulated for both types of deformations. Joint parameters and displacements of a new model are created just by applying the interpolators on a template model with new measurements. In Seo (2003) instead of statistically analyzing the anthropometric data, direct use of captured sizes and shapes of real people from range scanners is used in order to determine the shape in relation to the given measurements.

In former published papers, avatar motion models were based on simplified human skeleton with joints. In the early nineties, one of the first challenging methods for creating the human skeleton was published in Magnenat (1991), based on previous research for the hand skeleton (Gourret, 1989). They observed that existing avatar skeletons were more suitable for robots than for humans, thus a new skeleton layer was proposed. The trend was continued in (Monheit, 1991) with emphasis on providing more realistic effects for the torso that could be bent or twisted and (Scheepers, 1996) for the forearms and hands pronation and supination. A more recent model is detailed in Savenko (1999), based on initial investigations reported in Van (1998; 1999) where the focus is on improving the joint model and especially the knee kinematics. Performing realistic deformation was achieved by adding new layers in addition to the skeleton, namely muscle, fatty tissue, skin and clothing (Waters, 1989; Chadwick 1989; Scheepers, 1996; Singh, 1995). In Scheepers (1997) and Wilhelms (1997), the muscle layer is linked to the skeleton and is based on the anatomy of skeletal muscles. In Chen (1992), a finite-element model is presented, able to simulate the force of few individual muscles. In (Singh, 1995), a skin layer is attached to the skeleton layer, thus more local effects become visible.

Once the virtual character has been created, one should be able to change its postures in order to obtain the desired animation effect. The following section addresses the problem of virtual character animation and presents the main approaches reported in the literature.

Avatar animation

Animating a virtual character consists in applying deformations at the skin level. The major 3D mesh deformation approaches can be classified into the following five categories:

- Lattice-based (Maestri, 1999). A lattice is a set of control points, forming a 3D grid, which the user modifies in order to control a 3D deformation. Points falling inside the grid are mapped from the unmodified lattice to the modified one using smooth interpolation.
- Cluster-based (Maestri, 1999). Grouping some vertices of the skin into clusters enables to control their displacements by using the same parameters.
- Spline-based (Bartels, 1987). Spline and, in general, curve-based deformations allow deforming a mesh with respect to the deformation of the curve.
- Morphing-based (Blanz, 1999). The morphing technique consists in smoothly changing a shape into another one. Let us mention that such a technique is very popular for animating virtual human faces from pre-recorded face expressions.
- Skeleton-based. (Lander, 1999). The skeleton is a hierarchical structure and the deformation properties can be defined for each element of this structure.

The first four categories are used in animating specific objects, such as eyes (lattice), and facial expressions (morphing), and are more or less supported by the main animation software packages. The last category, more and more encountered in virtual character animation systems, introduces the concept of skeleton.

To design the virtual character skeleton, an initialization stage is necessary: the designer has to specify the influence region of each bone of the skeleton as well as a measure of influence. This stage is mostly interactive and recursively repeated until the desired animation effects are reached. When the skeleton moves, the new position of the vertex is calculated by multiplying the old position with the weights and matrices of the parent bones. While simple and easy to implement, the technique has some limitations, especially when animating soft body (the elbow problem). To overcome these problems, the basic technique was extended by different researchers. Lewis (2000) presented a solution in which they use different poses for the extreme situations where the skeleton animation failed. They save the information of these poses and associate it with the bone. This technique was improved by Kry (2002) by using Principal Component Analysis (PCA) to construct an error-optimal Eigen displacement basis for representing the potentially large set of pose corrections. The calculation is not done on the entire surface, but it is separated on more influence domains, thus optimizing it for use graphics hardware. [Wang02] proposed an alternative solution: instead of using one weight for each bone, weights are used for each component of the bone matrix. The weighting is done in a process in which the character is first animated, and then the weights are adjusted only for the problematic poses. [Mohr03] proposed a technique that improves the skeleton driven deformation by automatically adding new joints between existing ones to solve the problems in the extreme poses.

Since skeleton-based is the most used deformation model for avatars, we describe in the remaining part of the section what the approaches for animating the skeleton are. They can be classified into two categories: computer generated (kinematic, dynamic) and motion capture-based. A summary is illustrated in Table 1.

The kinematic approaches take into account critical parameters related to the virtual character properties such as position, orientation, and velocity. One of the classic solutions is to directly control the relative geometric transformation of each bone of the skeleton. This approach, also called Forward Kinematics (FK), is a very useful tool for the designer (Watt 1992) to manipulate the postures of virtual characters. The animation parameters correspond to the geometric transformation applied to each bone of the skeleton. An alternative approach is to fix the location in the world coordinates for a specific level of the skeleton, so-called end-effector (e.g. the hand for a human avatar), and to adjust accordingly the geometric transformation of its parents in the skeleton. With this method, also called Inverse Kinematics (IK), the animation parameters correspond to the geometric location of the end-effector.

Dynamic approaches refer to physical properties of the 3D virtual object, such as mass or inertia, and specify how the external and internal forces interact with the object. Such physics-based attributes have been introduced since 1985 in the case of virtual human-like models (Armstrong, 1985; Wilhelms, 1985). Extensive studies (Badler, 1995; Boston, 1998] on human-like virtual actor dynamics and control models for specific motions (walking, running, jumping, etc.) (Pandy, 1990 & 1999; Wooten, 1998) have been carried out. Faloutsos et al. (2001) proposed a framework making it possible to exchange controllers (i.e. a set of parameters) to drive a dynamic simulation of the character. The controller evolution is obtained by using the goals of the animation as an objective function. The results are physically plausible motions. Even if some positive steps have been achieved for specific motions, to simulate dynamically articulated characters displaying a wide range of motor-skills is still a challenging issue.

The motion capture technique (Menache, 2000) consists of tracking and recording the position (and the orientation) of a set of markers placed on the surface of a real object. Usually, the markers are positioned at the joints. The markers' positions, expressed in the world coordinate system, are then converted into a set of geometric transformations for each joint (Badler, 1993; Hirose, 1998; Molet, 99).

Motion capture technologies are generally classified into active and passive sensor-based capture according to the nature of the sensors used. With an active sensor-based system, the signals to be processed are transmitted by the sensors, while, in a passive sensor-based system, they are acquired by light reflection on the sensors. With respect to the nature of the sensors, the active sensor-based systems can be one of the following: mechanic-, acoustic-, magnetic-, optic- and inertial-based.

One of the earliest methods, using active mechanical sensors (Faro) is a prosthetic system. This is a set of armatures attached all over the performer's body and connected with a series of rotation and linear encoders. Reading the status of all the encoders allows for the analysis of the performer's postures.

The so-called acoustic method (S20sd) is based on a set of sound transmitters attached to the performer's body. They are sequentially triggered to emit a signal and the distances between transmitters and receivers are computed from the time needed for the sound to reach the receivers. The 3D position of the transmitter, and implicitly of the performer's segment, is then computed by using triangulation procedures or phase information.

Systems based on magnetic fields (Ascension, Polhemus) are made of one transmitter and several magnetic-sensitive receivers attached to the performer's body. The magnetic field intensity is measured by the receivers so that the location and orientation of each receiver are computed.

More complex active sensors are based on fiber optics. The principle consists in the measure of the light intensity passing through the flexed fiber optics. Such systems are usually used to equip data-gloves as proposed by VPL.[1]

The last method using active sensors is based on inertial devices, such as accelerometers, small devices which measure the acceleration of the body part to which they are attached (Aminian, 1998).

When using active sensors, the performer is burdened with a lot of cables, limiting his motion freedom. In this context, the recent developments of motion capture systems using wireless communication are very promising (Ascension, Polhemus).

The second class of motion capture techniques uses passive sensors. One camera, coupled to a set of mirrors properly oriented, or several cameras allow for the 3D posture reconstruction from these multiple 2D views. To reduce the complexity of the analysis, markers (light reflective or LEDs) are attached to the performer's body. The markers are detected on each camera view and the 3D position of each marker is computed. However, occlusions due to the performer's motions may get in the way. Additional cameras are generally used in order to reduce the loss of information and ambiguities.

Since 1995, computer vision-based motion capture has become an increasingly challenging issue when dealing with the tracking, posture computation and gesture recognition problems in the framework of human motion capture. The techniques can use only one image or sequence of images. Moeslund (2006) classifies them in three main branches: model-free, indirect model use, and direct model use. Model-free methods have no previous knowledge of the model so they take a bottom-up approach to track

[1] VPL Research Inc. Dataglove Model 2 Operation Manual, January 1989.

and label body parts in 2D. Indirect model methods use look-up a table to guide the interpretation of measured data. Direct model methods use previous knowledge of the model and try to use the data to find the model on the image. Included here are learning based methods that use training of the system with known poses (Agarwal, 2006).

Table 1. Summary of main animation techniques for avatars.

Computer Generated	Kinematic	Forward Kinematic
		Inverse Kinematic
	Dynamic	
Motion Capture	Active sensors	Mechanic
		Magnetic
		Optic
		Inertial
	Passive sensors	Reflective markers
		LEDs
		Computer vision

Standards, Recommendations and Markup Languages Related to Avatars

Other than the research community, the avatars have interested different standardization groups mainly due to the huge potential of applications involving them.2 There are currently two types of such standards: the ones interested in the appearance and the animation of the avatar in the 3D graphics applications (the avatars as representation objects) and the ones interested in avatars characteristics such as personality and emotions (the avatars as agents). In addition, there are several proprietary formats, imposed as de facto standards by the authoring tools or virtual world providers. All this multitude of solutions makes it impossible today to imagine even the simple scenario of using a single avatar for visiting two different virtual worlds. In this section we briefly introduce some of the standards from each category and give the main motivation behind MPEG-V.

Avatar Representation Standards

In the last decade, several efforts have been made to develop a unique data format for 3D graphics. In the category of open standards, X3D (based on VRML) and COLLADA are the best known, the latter being probably the most adopted by current tools. While COLLADA concentrates on representing 3D objects or scenes, X3D pushes the standardization further by addressing user interaction as well. This is performed thanks to an event model in which scripts, possibly external to the file containing the 3D scene, may be used to control the behavior of its objects. While the avatars in VRML/X3D are defined as specific objects being standardized under the name of H-Anim,3 in COLLADA there is no distinction between a human avatar and a generic skinned model. Also in the category of open standards, but specifically treating the compression of media objects, there is MPEG-4. Built on top of VRML, MPEG-4 contained, already in its first two versions (ISO, 1999), tools for the compression and streaming of 3D graphics assets, enabling to describe compactly the geometry and appearance of generic, but static objects, and also the animation of human-like characters. Since then, MPEG has kept working on improving its 3D graphics compression toolset and published two editions of MPEG-4 Part 16, AFX (Animation Framework eXtension) (ISO, 2004), which addresses the requirements above within a unified and generic

[2] Gartner Says 80 Percent of Active Internet Users Will Have A "Second Life" in the Virtual World by the End of 2011, http://www.gartner.com/it/page.jsp?id=503861.

[3] H-Anim – Humanoid AnimationWorking Group

framework and provides many more tools to compress more efficiently more generic textured, animated 3D objects. In particular, AFX contains several technologies for the efficient streaming of compressed multi-textured polygonal 3D meshes that can be easily and flexibly animated thanks to the BBA (Bone-Based Animation) toolset, making it possible to represent and animate all kinds of avatars.

While offering a full set of features allowing to display the avatars, none of the above-mentioned standards includes semantic data related to the avatar.

Agent-Related Standards and Recommendations

Several recommendations, standards or markup languages are related to adding semantics on top of virtual characters, mainly to describe features that do not necessarily have a visual representation (such as personality or emotions) or to expose properties that may be used by an agent (language skills, communication modality). The Human Markup Language (HumanML) (Brooks, 2002) by Oasis Web Services is an attempt to codify the characteristics that define human physical description, emotion, action, and culture through the mechanisms of XML, RDF and other appropriate schemas. HumanML is intended to provide a basic framework for a number of endeavors, including (but, as with human existence itself, hardly limited to) the creation of standardized profiling systems for various applications. It builds a framework for describing emotional state and response of both people and avatars, laying the foundation for the interpretation of gestures for both person-to-person and person-to-computer interpretations, the encoding of gestures and expressions to facilitate the better understanding of modes of communication.

EmotionML (EML) by W3C covers three classes of applications: manual annotation of material involving emotionality, such as annotation of videos, of speech recordings, of faces, of texts, etc; automatic recognition of emotions from sensors, including physiological sensors, speech recordings, facial expressions, etc., as well as from multi-modal combinations of sensors; generation of emotion-related system responses, which may involve reasoning about the emotional implications of events, emotional prosody in synthetic speech, facial expressions and gestures of embodied agents or robots, the choice of music and colors of lighting in a room, etc.

Behavior Markup Language (BML) (Vilhjalmsson, 2007) is an XML based language that can be embedded in a larger XML message or document simply by starting a <bml> block and filling it with behaviors that should be realized by an animated agent. The possible behavior elements include coordination of speech, gesture, gaze, head, body, torso face, legs, lips movement, and a wait behavior.

Multimodal Presentation Markup Language (MPML) (Ishizuka, 2000) is a script language that facilitates the creation and distributing of multimodal contents with character presenter. It also supports media synchronization with character agents’ actions and voice commands that conforms to SMIL specification.

Virtual Human Markup Language (VHML) is designed to accommodate the various aspects of Human-Computer Interaction with regards to Facial Animation, Body Animation, Dialogue Manager interaction, Text to Speech production, Emotional Representation plus Hyper and Multi Media information.

Character Mark-up Language (CML) (Arafa, 2003) is an XML-based character attribute definition and animation scripting language designed to aid in the rapid incorporation of life-like characters/agents into online applications or virtual worlds. This multi-modal scripting language is designed to be easily understandable by human animators and easily generated by a software process such as software agents. CML is constructed based jointly on motion and multi-modal capabilities of virtual life-like figures.

Avatar Representation and Semantics: The MPEG-V Vision

Despite the fact that several markup languages related to avatars and virtual agents exist, ensuring interoperability for avatars between different virtual worlds cannot be yet obtained in an easy, ready to use and integrated manner. Identifying this gap and recognizing that only the existence of a standardized format can make virtual worlds be deployed at very large scale, MPEG initiated in 2008 a new project called MPEG-V (Information exchange with virtual worlds). Concerning the avatars, the following requirements should be fulfilled by MPEG-V:

1) it should be possible to easily create importers/exporters from various VEs implementations,
2) it should be easy to control an avatar within an VE,
3) it should be possible to modify a local template of the avatar by using data contained in an MPEG-V file.

In the MPEG-V vision once the avatar is created (possibly by an authoring tool independent of any VW), it can be used in Second Life, in IMVU or in any other VW. An user can have his own unique presentation inside all VW, like in real life. He can change, upgrade, teach his avatar, i.e. "virtual himself" in one VW and then all the new properties will be available in all others. The avatar itself should then contain representation and animation features but also higher level semantic information. However, a VW will have its own internal structure for handling avatars. MPEG-V is not imposing any specific constraints on the internal structure of representing data by the VW, but only proposing a descriptive format able to drive the transformation of a template or a creation from scratch of an avatar compliant with the VW. All the characteristics of the avatar (including the associated motion) can be exported from a VW into MPEG-V and then imported in another VW. In the case of interface between virtual worlds and the real world (requirement 2), the avatar motions can be created in the virtual world and can be mapped on a real robot for the use in dangerous areas, for maintenance tasks or the support for disabled of elderly people and the like.

While the goal of MPEG-V is to obtain a descriptive format specifying the avatar features, it may be combined with MPEG-4 Part 16 (that includes a framework for defining and animating avatars) to provide a full interoperable solution.

Defining an interoperable schema as intended by MPEG-V can be of huge economic value being one step in the transformation of current VW from stand-alone and independent applications into an interconnected communication system, similar with current Internet where a browser can interpret and present the content of any web site. At that moment the VW providers will not be anymore providers of technology, but will concentrate their efforts on creating content, once again the success key of Internet.

MPEG-V Schema description

Based on an analysis of existent VWs and most popular games, tools and techniques from content authoring packages, together with the study of different virtual human-related markup languages, the current version of MPEG-V[4] defines a set of metadata referring to the appearance, animation, and agent-like capabilities of an avatar. In this section we are presenting the main elements of the schema.

The "Avatar" element is composed of following type of data (for a detailed explanation and for exact schema definition, please refer to Preda (2009).

[4] MPEG-V was promoted as Comity Draft in July 2009.

Appearance, Animation and Haptic Properties

The Appearance element contains descriptions of the avatar's different anatomic segments (size, form, anthropometric parameters) as well as references to the geometry and texture resources. While the first can be used to adapt the internal structure of the VW avatar (personalizing it), the second can be used to completely overwrite it (operation performed when the format for the resource itself is also known by the importer/exporter—such as the case when using MPEG-4 3D Graphics). In addition, this element also contains characteristics of objects that are related to the avatar such as clothes, shoes, or weapons. A simple and very short example of how this elements is used in MPEG-V is given below:

```
<Appearance>
   <Body > <BodyHeight value=165 /> <BodyFat value=15 /> </Body >
   <Head> <HeadShape value="oval" /> <EggHead value="true" /> </Head>
   <Clothes ID=1 Name="blouse_red" />
   <AppearanceResources> <AvatarURL value="my_mesh" />
   </AppearanceResources>
</Appearance>
```

The Animation element contains a complete set of animations that the avatar is able to perform, grouped by semantic similarity (Idle, Greeting, Dance, Walk, Fighting, Actions). A special group contains common actions such as drink, eat, talk, read, and sit. As in the previous case, the animation parameters are represented in external resources, MPEG-V providing only the names of the animation sequences. A simple example of using this element is given below:

```
<Animation>
   <Greeting > <hello> my_hello.bba </hello> <wave>my_wave.bba </wave></ Greeting >
   <Fighting> <shoot>my_shoot.bba</shoot> <throw>my_trhrow.xml</throw> </ Fighting >
   <Common_Actions> <drink>my_drink.bba</drink> <eat>my_eat.xml </eat>
<type>my_type.xml</type> <write> my_write.xml </write></ Common_Actions>
   <AnimationResources> <AvatarURL value="my_anim”/> </AnimationResources>
</Animation>
```

The Haptic properties are defined with the main purpose of simulating feed-back from VWs. If "haptic gloves" or "tactile screen" are used, the touch can be rendered as vibrations or force field.

These first three elements are used in MPEG-V for ensuring portability of the avatar graphic representation between different VW.

Control

The main purpose of this element is to provide the correspondence between the avatar control parameters in a VW (bones of the skeleton, feature points on the face mesh) and a standardized set of controllers (defined as an exhaustive list of bones and feature points). When an input signal (i.e. the position and orientation of a magnetic sensor) is connected to one controller, the mapping between the latter and the avatar's bone allows its animation. Knowing the correspondence between the generic skeleton and the one of the specific avatar in the VW makes it possible to map entire animation sequences (motion retargeting). A simple example is provided below:

```
<Control>
   <BodyFeaturesControl >
      <headBones>
         <CervicalVerbae3>my_cervical_verbae_3</CervicalVerbae3>
         <CervicalVerbae1> my_cervical_verbae_1</CervicalVerbae1>
         <skull>my_skull</skull>
     </headBones>
     <UpperBodyBones>
         <LCalvicle>my_ LCalvicle</LCalvicle>
         <RClavicle>my RCalvicle</RClavicle>
      </UpperBodyBones>
    </BodyFeaturesControl >
   <FaceFeaturesControl >
      <HeadOutline>
         <Left X=0.23 Y=1.25 Z=7.26 />
         <Right X=0.25 Y=1.25 Z=7.21 />
         <Top X=2.5 Y=3.1 Z=4.2 />
         <Bottom X=0.2 Y=3.1 Z=4.1 />
      </ HeadOutline >
   </FaceFeaturesControl >
</ Control>
```

This element ensures controlling the avatar from external signals and motion retargeting.

Communication Skills and Personality

CommunicationSkills (Oyarzun, 2009) defines the way that avatar is able—or wants—to communicate with other avatars. The way of communicating is characterized by the input and output abilities, both for verbal and non-verbal communication. It contains sub-elements that describe language, verbal or sign, that the avatar can interpret/understand, preferred mode of communication, preferred language etc. A simple example is given below:

```
<CommunicationSkillsType >
   <InputVerbalCommunication Voice="prefered" Text="enabled" >
      <Language Name="French" Preference="Text" />
      <Language Name="English" Preference="Voice" />
      <Language Name="Italian" Preference="Voice" />
   </InputVerbalCommunication>
   <OutputVerbalCommunication Voice="prefered" Text="enabled" >
      <Language Name="French" Preference="Voice" />
   </OutputVerbalCommunication>
</CommunicationSkillsType>
```

Personality (Oyarzun, 2009) is described as a combination of openness, conscientiousness, extraversion, agreeableness, and neuroticism as defined in the OCEAN model (McCrae, 1992). Personality can serve to adapt the avatar's verbal and non-verbal communication style as well as modulating emotions and moods that could be provoked by virtual world events, avatar-avatar communication or the real time flow.

At the time of writing this paper, MPEG-V is still in progress. Near future perspective is related to representing the avatar emotions. The final version of MPEG-V is planned for the end of 2010.

Conclusion

The advancements achieved in the last two decades in the field of avatars with respect to their visual appearance (static and animating) as well as cognitive properties make it possible to imagine today an integrated representation solution aiming to ensure one of the main requirements of interoperability between virtual worlds: being able to migrate from one world to another while maintaining the user properties. These are encapsulated together in what is commonly called today the user avatar. It is expected that each future VW will use sub-sets of user properties; probably all of them will use appearance and animation properties to ensure the visual representation. Some properties/capabilities obtained in one VW remain connected to the avatar and should be available as well to the outside worlds (real or virtual). Providing the container of such properties/capabilities is the vision of MPEG-V. By its descriptive format it aims to facilitate the deployment of VW recognizing that their success depends on maintaining acceptable development cost and one mechanism in doing so consists in ensuring interoperability between them at least at the level of avatars.

Bibliography

Agarwal, A. And Triggs, B. 2006. (2006, January). Recovering 3D Human Pose From Monocular Images. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1, 44-58.

Aminian, K., Andres, E. D., Rezakhanlou, K., Fritsch, C., Schutz, Y., Depairon, M., Leyvraz, P., And Robert, P. (1998). Motion Analysis In Clinical Practice Using Ambulatory Accelerometry. In Proceedings Of The International Workshop On Modelling And Motion Capture Techniques For Virtual Environments N. Magnenat-Thalmann And D. Thalmann, Eds. Lecture Notes In Computer Science, Vol. 1537. Springer-Verlag, London, 1-11.

Arafa, Y., And Mamdani, A. (2003). Scripting Embodied Agents Behaviour with CML: Character Markup Language. In IUI ’03: Proceedings of the Eigth International Conference on Intelligent User Interfaces, ACM, New York, NY, USA, 313–316.

Armstrong, William W. And Green, Mark W. (1985). "The Dynamics of Articulated Rigid Bodies for Purposes of Animation". Proc. Graphics Lnterfafe 85, 407-415.

Ascension Technology Motionstar® Http://Www.Ascensiontech.Com/Products/Motionstar/.

Badler N., Metaxas D., Webber B. And Steedman M. (1995). The Center for Human Modeling and Simulation, Presence 4, 1, 81-96.

Badler N., Phillips C., and Webber B. (1993). Simulating Humans: Computer Graphics, Animation, and Control. Oxford University Press.

Bartels, R. H., Beatty, J. C., And Barsky, B. A. (1987). An Introduction to Splines for Use in Computer Graphics & Amp; Geometric Modeling. Morgan Kaufmann Publishers Inc.

Blanz, V., Vetter, T. (1999). A Morphable Model For The Synthesis Of 3D Faces. Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, 187-194.

Boston Dynamics Inc. (1998). The Digital Biomechanics Laboratory, Www.Bdi.Com.

Brooks, R., And Cagle, K. (2002). The Web Services Component Model And Humanml. Technical Report, OASIS/Humanml Technical Committee.

Chadwick, J. E., Haumann, D. R., and Parent, R. E. (1989, July). Layered Construction for Deformable Animated Characters. Computer Graphics (SIGGRAPH 89 Conference Proceedings) 23, 3, 243–252.

Chen, D. T. And Zeltzer, D. (1992, July). Pump It Up: Computer Animation of a Biomechanically Based Model of Muscle Using the Finite Element Method. SIGGRAPH Comput. Graph. 26, 2, 89-98.

D’Apuzzo, N., Plankers, R., Fua, P., Gruen, A., Thalmann, D.: Modeling Human Bodies From Video Sequences. Videometrics VI, SPIE Proceedings, Vol. 3461, San Jose, CA, 36-47, 1999.

Decarlo, D., Metaxas, D., And Stone, M. (1998). An Anthropometric Face Model Using Variational Techniques. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '98.

Devernay, F. And Faugeras, O. D. (1994). Computing Differential Properties Of 3-D Shapes from Stereoscopic Images without 3-D Models. In Conference on Computer Vision and Pattern Recognition, Seattle, WA, 208-213.

Dooley M. (1982). Anthropometric Modeling Programs –A Survey”, IEEE Computer Graphics And Applications, IEEE Computer Society, 2, 9, 17-25.

EML, Emotionml, Http://Www.W3.Org/2005/Incubator/Emotion/XGR-Emotionml/.

Faloutsos, P., Van De Panne, M., And Terzopoulos, D. (2001). Composable Controllers For Physics-Based Character Animation. In Proceedings Of The 28th Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '01. ACM, New York, NY, 251-260.

Faro Technologies, http://www.Farotechnologies.Com.

Georges, F. (2009). Représentation de soi et identité numérique: analyse sémiotique et quantitative de l’emprise culturelle du web 2.0. In Réseaux: Usages du Web 2.O, 27, 154. Paris: La découverte.

Gourret, J.-P., Thalmann, N. M., And Thalmann, D. (1989, July). Simulation Of Object And Human Skin Deformations In Agrasping Task. Computer Graphics (SIGGRAPH 89 Conference Proceedings) 23 (July), 21–30

Hilton, A., Beresford, D., Gentils, T., Smith, R., And Sun, W. 1999. Virtual People: Capturing Human Models To Populate Virtual Worlds. In Proceedings Of The Computer Animation (May 26 - 28, 1999). CA. IEEE Computer Society, Washington, DC, 174.

Hirose, M., Deffaux, G., & Nakagaki, Y. (1996). Development Of An Effective Motion Capture System Based On Data Fusion And Minimal Use Of Sensors. VRST'96, ACM-SIGGRAPH And ACM-SIGCHI, 117-123.

Ishizuka, M., Tsutsui, T., Saeyor, S., Dohi, H., Zong, Y.,And Predinger, H. (2000). Mpml: A Multimodal Presentation Markup Language with Character Agent Control Functions. Web-Net.

ISO/IEC JTC1/SC29/WG11. (2004). Standard 14496-16, A.K.A. MPEG-4 Part 16: Animation Framework Extension (AFX), ISO.

ISO/IEC JTC1/SC29/WG11. (1999). Standard 14496-2, A.K.A. MPEG-4 Part 2: Visual, ISO.

Kry, P. G., James, D. L., And Pai, D. K. (2002). Eigenskin: Real Time Large Deformation Character Skinning In Hardware. In *Proceedings Of The 2002 ACM Siggraph/Eurographics Symposium On Computer Animation* (San Antonio, Texas, July 21 - 22, 2002). SCA '02. ACM, New York, NY, 153-159

Lander J. (1999, May) Over My Dead, Polygonal Body. Game Developer Magazine, 1--4.

Lewis, J. P., Cordner, M., and Fong, N. (2000). Pose Space Deformation: A Unified Approach to Shape Interpolation and Skeleton-Driven Deformation. In *Proceedings of The 27th Annual Conference on Computer Graphics and Interactive Techniques* International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, 165-172.

Maestri G. (1999, July). Digital Character Animation 2: Essential Techniques. New Riders

Magnenat-Thalmann, N., and Thalmann, D. (1991, September). Complex Models For Animating Synthetic Actors, *IEEE Computer Graphics And Applications 11*, 5, 32–44.

Marcus G. Pandy and Frank C. Anderson. (1999, August). Three-Dimensional Computer Simulation of Jumping and Walking Using the Same Model. In *Proceedings of the Seventh International Symposium On Computer Simulation In Biomechanics*.

Menache, A. (1999). Understanding Motion Capture for Computer Animation and Video Games. 1st. Morgan Kaufmann Publishers Inc.

Moeslund, T. B., Hilton, A., and Krüger, V. (2006, November). A Survey of Advances in Vision-Based Human Motion Capture and Analysis. *Comput. Vis. Image Underst. 104*, 2, 90-126.

Mohr, A. and Gleicher, M. (2003, July). Building Efficient, Accurate Character Skins from Examples. *ACM Trans. Graph. 22*, 3, 562-568.

Molet, T., Boulic, R., and Thalmann, D. (1999, April). Human Motion Capture Driven By Orientation Measurements. *Presence: Teleoper. Virtual Environ. 8*, 2, 187-203.

Monheit, G., and Badler, N. I. (1991, March). A Kinematic Model of the Human Spine and Torso, *IEEE Computer Graphics and Applications 11*, 2, 29–38.

NASA. (1995, July). Man-Systems Integration Standard (NASA-STD-3000), Revision B.

NASA. (1978). Reference Publication 1024, The Anthropometry Source Book, Volumes I And II.

Nebel J., Sibiryakov A. (2002). Range Flow from Stereo-Temporal Matching: Application to Skinning, In: *Proceedings Of IASTED International Conference On Visualization, Imaging, And Image Processing*.

Oyarzun, D., Ortiz, A., del Puy Carretero, M., Gelissen, J., Garcia-Alonso, A., and Sivan, Y. (2009). ADML: A framework for Representing Inhabitants in 3D Virtual Worlds. In Proceedings of the 14th international Conference on 3D Web Technology (Darmstadt, Germany, June 16 - 17, 2009). S. N. Spencer, Ed. Web3D '09. ACM, New York, NY, 83-90.

Pandy, M. G., Zajac, F. E. (1990). An Optimal Control Model For Maximum-Height Human Jumping. Journal Of Biomechanics, 23(12):1185–1198,

Preda M. (Ed.). (2009). Text of ISO/IEC CD 23005-4 Avatar Information, w10786, 89th MPEG Meeting, London.

Polhemus STAR*TRACK Motion Capture System, Http://Www.Polhemus.Com

S20SD S20 Sonic Digitizers, Science Accessories Corporation.

Savenko A., Van Sint Jan S.L. and Clapworthy G.J. (1999). A Biomechanics-Based Model for the Animation of Human Locomotion, *Proc Graphicon 99*, Moscow, 82-87.

Scheepers, C. F. (1996). *Anatomy-Based Surface Generation for Articulated Models of Human Figures*. Phd Thesis, Ohio State University, Adviser: Richard E. Parent.

Scheepers, F., Parent, R. E., Carlson, W. E., and May, S. F. (1997). Anatomy-Based Modeling of the Human Musculature. In *Proceedings Of The 24th Annual Conference on Computer Graphics and Interactive Techniques.* International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, 163-172.

Scheepers, F., Parent, R. E., May, S. F., and Carlson, W. E. (1996, January). A Procedural Approach To Modelling And Animating The Skeletal Support Of The Upper Limb, Tech. Rep. OSUACCAD-1/96-TR1, ACCAD, The Ohio State University.

Seo, H. and Magnenat-Thalmann, N. (2003). An Automatic Modeling of Human Bodies from Sizing Parameters. In *Proceedings of the 2003 Symposium on Interactive 3D Graphics* (Monterey, California, April 27 - 30, 2003). I3D '03. ACM, New York, NY, 19-26.

Seo, H., Yahia-Cherif, L., Goto, T., and Magnenat-Thalmann, N. (2002). GENESIS: Generation of E-Population Based on Statistical Information. In *Proceedings of the Computer Animation* (June 19 - 21, 2002). CA. IEEE Computer Society, Washington, DC, 81.

Singh, K. (1995). *Realistic Human Figure Synthesis And Animation For VR Applications*. Phd Thesis, The Ohio State University. Adviser: Richard E. Parent

Van Sint Jan S.L., Salvia P., Clapworthy G.J., Rooze M. (1999). Joint-Motion Visualisation Using Both Medical Imaging And 3D- Electrogoniometry, Proc 17th Congress Of International Society Of Biomechanics, Calgary (Canada).

Van Sint Jan, S. L., Clapworthy, G. J., And Rooze, M. (1998, November). Visualization of Combined Motions in Human Joints. *IEEE Comput. Graph. Appl. 18*, 6, 10-14.

VHML, Http://Www.Vhml.Org/.

Wang, X. C. And Phillips, C. (2002). Multi-Weight Enveloping: Least-Squares Approximation Techniques for Skin Animation. In *Proceedings of the 2002 ACM Siggraph/Eurographics Symposium on Computer Animation* (San Antonio, Texas, July 21 - 22, 2002). SCA '02. ACM, New York, NY, 129-138

Waters, K. (1989). Modeling 3D Facial Expressions: Tutorial Notes. In *State Of The Art In Facial Animation.* ACM SIGGRAPH, 127–160.

Watt, A. And Watt, M. (1991). *Advanced Animation And Rendering Techniques*. ACM.

Wilhelms, J. And Van Gelder, A. (1997). Anatomically Based Modeling. In *Proceedings of The 24th Annual Conference On Computer Graphics And Interactive Techniques* International Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, 173-180.

Wilhelms, J. P. And Barsky, B. A. (1985). Using Dynamic Analysis to Animate Articulated Bodies Such as Humans and Robots. In *Proceedings Of Graphics Interface '85 on Computer-Generated Images: The State of the Art* (Montreal, Quebec, Canada). N. Magnenat-Thalmann And D. Thalmann, Eds. Springer-Verlag New York, New York, NY, 209-229.

Wooten, W. L. (1998). *Simulation Of Leaping, Tumbling, Landing, and Balancing Humans*. Doctoral Thesis. UMI Order Number: AAI9827367, Georgia Institute Of Technology.

3dsmax, 3D Studio Max, Autodesk. Webpage: Http://Www.Autodesk.Com/3dsmax

Maya, Autodesk. Webpage: Http://Www.Autodesk.Com/Maya

Ronfard R. Ronfard, J. and J. Rossignac. (1996, August). Full-Range Approximation of Triangulated Polyhedra, Proceedings EUROGRAPIHCS, Computer Graphics Forum, 67-76.

Schroeder, W. J., Zarge, J. A., and Lorensen, W. E. (1992). Decimation of Triangle Meshes. In *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques* J. J. Thomas, Ed. SIGGRAPH '92. ACM, New York, NY, 65-70.

Vilhjalmsson, H. and Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z.M., Thórisson, K., van Welbergen, H. and van der Werf, R.J. (2007, September). *The Behavior Markup Language: Recent Developments and Challenges.* In: Proceedings of the Seventh International Conference on Intelligent Virtual Agents, Paris, France.