Simple emphatic user interface for virtual heritage

Stanek Stanislav

Comenius University
Mlynská dolina
84248, Bratislava, Slovakia

stanek@sccg.sk

ABSTRACT

We present a simple user interface combining h-Anim with Perlin's face. The main application is for exploring virtual environments especially those representing real environments with places, buildings or objects that belong to cultural heritage or those with historical past, famous story or something interesting. Therefore, information about them should be delivered to user. This kind of information is usually full of emotions and that is why the most suitable way (from user interface point of view) is to deliver it with emphatic storytelling. We are introducing our simple emphatic system (implementation uses ActiveX objects, VRML, ECMA Script, Java Script) that uses simple hardware configuration with web cams used for capturing user's presence and his/her head movements and if possible capturing position of some facial features, defined in MPEG-4 standard, and used to recognize user's simple emotions. User presence, head movements, and simple emotions are used to create simple emphatic user interface. In this paper we present our results already used in some application projects for virtual museums.

Keywords

Virtual environments, user interface, storytelling, emphatic, autonomous agents, capture, interaction

1. INTRODUCTION

We present a simple user interface combining h-Anim with Perlin's face. The main application is for exploring virtual environments (VEs) especially those representing real environments with places, buildings or objects that belong to cultural heritage or those with history, famous story or something interesting. This information is also usually full of emotions and that is why the most suitable way (from user interface point of view) is to deliver it with emphatic storytelling to user or visitor.

In the real world it would be a real person that is telling a story, answering questions, expressing emotions that come with story and also creating emotions that are arising from a simple conversation with a visitor. Emphatic communication between a storyteller and a visitor is in real world set very naturally this way.

However, VEs [QVO01,QVO02] it is not so easy to set up this kind of emphatic communication. We are introducing our simple emphatic system that uses simple hardware configuration with web cams used for capturing user's presence and his/her head movements (position and orientation) and if possible capturing position of some of his/her facial features, defined in MPEG-4 standard, and used to recognize user's simple emotions. According captured information like user presence, head movements and simple emotions, our simple emphatic interface is creating actions and reactions (eg. starts/stops presentation) and introducing simple empathy this way. To achieve this emphatic feeling we are also using human-like autonomous agents as a part of this user interface. At this time we are working on middle precision autonomous agent based on minimal Perlin's face [PER00] structure and structure of its body is H-Anim 200x [HAnim] compatible with few extensions.

The rest of the paper is organized as follows. In background section some parts of recent project are described in more detail. We briefly describe previous work. Than we present our recent work and project in which recent results are applied. Finally, we outline future work.

2. BACKGROUND

Emphatic autonomous agents

Trying to construct empathic autonomous agents, we have to consider the immersion of a guest in VE and the immersion of the autonomous agent into the feelings of real visitors of VEs. Our idea is to limit the empathy of an avatar with respect to the well-known contexts. If the guest is very distant, no empathy makes sense. Strictly speaking, there are two sources of avatar's "emotionality". It is the message itself (sad story, ballad, funny paradox, irony…) and the guest's reactions. The first layer of emotions is ready in advance, say, off-line. We will use it for message presentation, e.g. while looking and pointing at the Crowning Tower of Bratislav Castle [VHCE04].

Another communication layer - guest's reactions - are available on-line only. They can be derived from the distance, head movements and facial expressions. This can be subdivided to data from user profile (ID, gender, etc. – data record) and guest behavior (navigation, exploration, interaction, cooperation, expressions, emotions …).

The two layers of emotionality are combined using the algebra of facial expressions. We will weight the importance of the on-line layer. To clarify this we give two examples. Both start when tourist guide explains a story of a castle.

1. When a guest rotates his/her head, then the guiding autonomous agent will nonverbally respond to this. She/he will add to her/his head movement a half-way angle rotation. For this we have a metaphor of a delayed and lazy mirror.

2. When a guest quits his/her tour, then the guiding autonomous agent has to cancel the original role by saying "Goodbye and See you soon here in our castle".

The difference between the two examples is in a different use of the signal from guest. In the first case the response is added, whereas in the second case the importance of a signal is absolute and there is nothing to combine.

We will limit the movement of the guest and guiding avatar. The inspiration arose from Virtual Old Prague project [ŽÁRA02], where the city sectors are separated by invisible walls. Their touching works as sensors for fetching of another part of VRML/MySQL database. This improves the real-time rendering budget. Our idea is to surround visible objects (autonomous agents included) by multiple glass walls. The closest of them will prevent unreal close-ups and thus limit the needs for texture storage and processing. The area between two glass walls surrounding the autonomous agent can be characterized in terms of rendering speed. The information tells to avatar which precision of empathy is practically needed. E.g. for a very distant user no facial expressions are needed. The middle distance signalizes the need for voice and simplified (Perlin) face. The nearest level of detail will start the complete use of MPEG-4 complaint full emphatic functionality.

The feasible combination of levels of details is under study (stories, photorealism, and autonomous agents empathy).

Modeling autonomous agents

The whole model of autonomous agent is hierarchical, segmented structure defined by joints and segments as specified in H-Anim standard representation for humanoids [HAnim]. So, the position and possible transformations per every segment using its joint are defined. For every segment there are defined also deformations that are used to create expressions of the segment.

Using Perlin's face we are able to create all basic facial expressions defined in MPEG-4 standard (anger, fear, sadness, surprise, joy and disgust) ([ABR99], Figure 1). Perlin's face uses minimal structure needed to create readable basic facial expressions. In the current work we are more concerned to create facial expressions and head movements in real-time for VE and that is why Perlin's minimal face is useful for us. It can be said that this face is defined, so to say, from approximate subset of feature points defined in MPEG-4 standard and that is why we refer to it as medium precision. We are planning for the future to extend this face to high precision face that will be compatible with MPEG-4 definitions and will contain all feature points defined in MPEG-4. Perlin's face structure in our application is transformed to satisfy H-Anim standard (consists of segments, joints, …) Facial expression defined in Perlin's face model is transformed to deformations of face segments in H-Anim representation. And body structure also satisfies H-Anim standard. So our model is created as H-Anim humanoid model.

Using H-Anim standard for body representation allows us to use also body language to extend facial expressions with whole body expressions and also allows us this way to extend facial emphatic communication with emphatic communication created with body movements (for example head movements or hand gestures). The structure and functionality of body parts can be also referred to as a medium precision, because in high precision also their deformations and more complex structure will be considered. Using this standard it is also helpful for creating simple motions and emotions. Also the application of pre-captured or real-time captured motion data is not complex. At this moment we are using pre-captured motions and expressions stored in a motions and expressions library.

VRML prototypes

We are concerned to VRML virtual environments and our autonomous agent is also VRML model and it uses H-Anim prototypes. We implemented in VRML its functionality (scripting with ECMA Script) using structure expressions that are created using other structure expressions or deform expressions or rotation expressions. Because model of autonomous agent is a hierarchical VRML structure consisting of segments that are substructures of the structure and any segment can have defined any expressions, we created functional prototypes for any type of expression. The prototypes are defined for reusability. We use these prototypes for export of created model to VRML file supporting defined functionality of the autonomous agent.

For our purposes we also extended H-Anim standard prototypes with some VRML nodes to achieve needed functionality. We introduced some extensions to Joint and Displacer prototypes that are parts of Humanoid prototype defined in H-Anim

As for the Segment node in H-Anim is defined Displacer node we introduced Rotator node for Joint node in H-Anim.

```
PROTO Joint [ …
  # Extended with:
  exposedField MFNode rotators [ ]
  exposedFieldSFNode rotationControler NULL
  exposedField MFNode StructureExpressions [ ]
  exposedField MFNode Sounds [ ]
]

PROTO Displacer [ …
  # Extended with:
  field SFNode Segment Segment { }
  field MFNode DeformExpressions []
  eventIn SFNode set_DeformExpressionState
]

PROTO Rotator [
  exposedField SFString name “”
  exposedField SFInt32 ID -1
  exposedField SFRotation rotation 0 0 1 0
  exposedField SFFloat weight 0
  exposedField MFFloat weight_ranges [-1 0 1]
  exposedField MFRotation orientations [ ]
  field MFNode RotationExpressions [ ]
  eventIn SFNode set_RotationExpressionState
…
]
```

Prototypes RotationExpressions are used to create body and head motions and emotions (body language) applying rotational transformations to joints of avatar’s body structure. Prototypes DeformExpressions are used to create facial expressions applying translation transformations to points of segments (segment deformations) of avatar’s body structure. For example, DeformExpressions are also used to create mouth movements corresponding to actual words of an autonomous agent. At least to each phoneme defined in MPEG-4 there is corresponding DeformExpression.

Sound node in Joint node is used to correctly localize voice of an autonomous agent (it has to come from its mouth and if the avatar is far away, you are not able to hear it, but if you get closer to it you will hear the words it is saying. This functionality comes from specification of VRML files and using audio in VRML environments.).

Because of reusability with setting only some parameters we also created new prototypes StoryTeller, Timeline, TimelineAction and ActionEvent. Using these prototypes we are defining functionality that is needed to tell a story to a user. That means that these prototypes are used to store what to say, how to say it and when to say it. They are also taking care about all synchronizations and especially they are taking care about synchronizations of voice with lips, facial and body expressions.

This way, having textual representation of the audio with story, we have information about the mouth deformation sequence in time. Additionally we have to synchronize this deformation sequence with storytelling audio. So storytelling timeline has defined marks for this synchronization. Definition of these marks is at this moment done manually with only a simple automatic processing, but this will become automatic with integration with system for facial feature capturing.

All prototypes are defined so that we are able to do simple combinations (or interpolations) of any expressions.

We created this "language" also because we need to combine different motions that can be divided into at least two layers of motions respectively expressions or emotions. The first layer of motions and emotions is defined with the content of the story. The second one is defined by the environment in which the storytelling avatar actually is. This second layer also consists of motions and expressions that are created according to given information about position in environment and for example also information about position of an object that it is actually speaking about. To the second layer belong also expressions and movements created as reactions to captured and recognized expressions of user if capturing system is working.

All defined prototypes are using advantages of TimeSensors and ROUTEs that are defined in VRML specification.

3. PREVIOUS WORK

Our work is still in progress but some achieved partial results are already used in some projects. Final results will be used in many other running projects presenting real environments using its virtual approximation and each of these projects needs storytelling in VE.

Older results are already used in international project [VHCE04] that is dealing with cultural heritage in different countries. Our autonomous agents are in this project used as emphatic virtual storytellers or emphatic tour guides, telling stories with emotions achieved with body and head motions, facial expressions and with speech synchronized lips movements.

We created simple user interface for VEs that offers to see predefined viewpoints and predefined guided tour to user. Our autonomous agent is a key object in this interface. On the figure 1 you can see an example of created user interface used in VHCE project and description of it. Using this interface user can switch between predefined viewpoints, call storyteller, and turn on and off its emotions (humanlike behavior) and also to view predefined guided tour. Predefined guided tour is made of motion trajectory used as guide for flying around, close to or inside of presented objects and during this fly the storyteller is telling interesting story. This guided tour can be stopped, paused or played again at any time.

Emotional layer of autonomous agent is created in two sub layers. First one is defined by the story and its emotions and that is why these emotions are the same each time the storyteller is telling the story. We use predefined body and head expressions defined by rotations and facial expressions with moving lips synchronized to speech defined by deformations and time stamps.

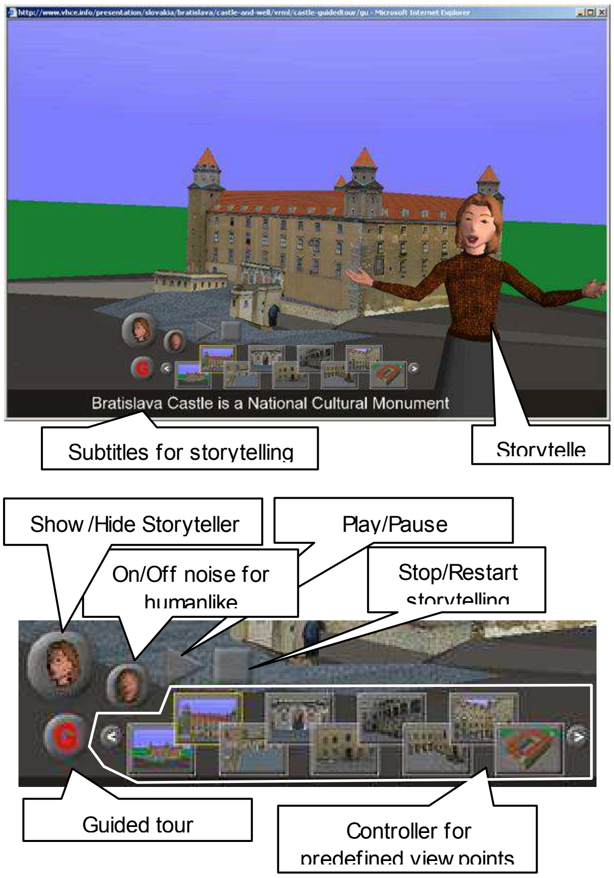

Figure 1. UI for VRML worlds with avatar (functionality description), Screenshot from project VHCE [VHCE].

Second layer is independent of the storytelling. It represents emotions according to situation in which the storyteller actually is (eg. position in VE or in the future also movements and emotions of visitor in VE). These emotions are created by applying facial expressions (deform expressions) and head and body expressions (rotational expressions) in time with weight that is processed with Perlin's noise and that is why it look so random and natural. This second layer helps us to create autonomous agent that acts approximately like a real human.

For user interface in VE we created also graphical and functional prototypes that can be used to create custom interfaces depending on application or type of VE. All these prototypes are also used for creating of user interface for prototype GuidedTour in VHCE project that has integrated our empathic autonomous agent. These prototypes are nowadays used and extended to fulfill functionality needed for virtual museums (see next section).

4. VIRTUAL MUSEUM

The latest results of our research are planned to be used and implemented in many project. We will describe our latest results by describing the project “Považské múzeum 3D” where these results are implemented.

Project description

This project is dealing with creating virtual representation of a real museum called “Považské múzeum“ [PMZ08]. For this project we created special hardware configuration or “hot spot” or as we are saying “well of knowledge” where people can virtually visit this museum and get a lot of information about this museum (see Figure 2).

This design for the “hot spot” in this project has triple metaphorisation - a drop of water falling up combined with the view through a locked door, using just the keyhole into the 13th chamber, where the source of knowledge is hidden.

Figure 2. Gateway to virtual museum or “hot spot” or “well of knowledge”. (photos by Ela Šikudová)

The third metaphor is inside - the source of living water. The virtual water has sound rendering, as well, and after double-click it metamorphoses into a book, symbolic source of knowledge. The drop falling up symbolizes the visit of a museum. Go back in the time

Figure 3. Design metaphors and screenshots from presentation layer.

and causality and refresh your memory. The UNESCO page mission is credited here "Heritage is our legacy from the past, what we live with today, and what we pass on to future generations." (see Figure 3)

Functionality

This "hot spot" is a gateway to virtual museum and is created for any kind of user and that is why the user interface should be as simple as possible. And also to be able to deliver information about this museum in a natural way we are introducing emphatic user interface achieved with built-in web cams that are capturing and monitoring situation close to display part. (see Figure 2.) In this hardware configuration according to captured space in front of the hot spot web cams have these three main functionalities:

1. Identify presence of a possible user
2. Identify head of a user and capture its movements
3. If possible identify facial features and try to identify users emotions

First two main functions are already working. First one is used to send presentation into a sleeping mode (like screensaver with black screen, in this project there is just a simulation of a well with water and with pure lighting) when there is no user detected (no motion is detected) or it will "wake up" presentation (starts displaying presentation on the screen, in this project it means that a simulation of a well with water and lighting is displayed) when there is a possible user detected.

Second one is used to identify that a person is present and that a book of knowledge should be rendered inside of the well under the water waiting for next interaction through touch display.

Touching the display is just simulating touching of water. This part is created with flash technology and ActionScript. After double-click on the top of the water it brings the book of knowledge in front of the water and user can choose from topics that he/she is interested in. Available topics are information about project partners and presentations about museum with simple user interface for choosing a presentation, to play and stopping, moving forward or back slides in presentation. Presentation and its slides are predefined and can consist of images and videos.

One of those presentations is created as a VE with 3D model of a castle where the museum is placed and our emphatic autonomous agent with predefined guided tour is ready to tell a simple story about this museum (Figure 4.). In this part second and third web cam functionalities are mainly used. Because we are using web cams with low resolution, nowadays we are not successful with automatic facial features extraction and that is why we have also problem with identification of any simple emotion of present user.

Figure 4. Guided tour. Virutal environment without textures.

System functionality

In this section we will simply describe functionality of our system.

There is a simple block diagram of our system on the figure 5. We are using a web browser (Internet Explorer or Firefox) for which ActiveX objects are created and used. It is because in the future this system should be placed and work on the Internet.

For VEs created in VRML [VRML97] we use viewer Cortona (www.parallelgraphics.com) as ActiveX component in browser. It is used as one of many viewers for VEs created in VRML. We use it because it is also viewer that is commonly used. We created VRML prototypes for our user interface between user and VE (see VRML prototypes). Guided tour that consists of predefined viewpoints and autonomous agent is part of this interface.

There is another ActiveX object in browser created by VideoForge system [KUB06] that is responsible for capturing and analyzing captured data from cameras. VideoForge is working in real time and it should have actual information about presence of a user, his/her head movements and in the future also movements of facial features of the user. It is using simple and complex filter sequences for detection and features tracking. We created an Interface as a communication bridge between VideoForge and Cortona. This interface is nowadays for our testing purposes just a simple set of JavaScript functions. VideoForge and Cortona are working separately. Interface is generating questions to Videoforge about user and answers to this questions are analyzed and the results are sent to Cortona where Guided tour respectively autonomous agent is responding to this results. Interface is also asking question to Guided tour and autonomous about actual viewing situation and depending on answer it is needed or not to send results or ask questions to VideoForge.

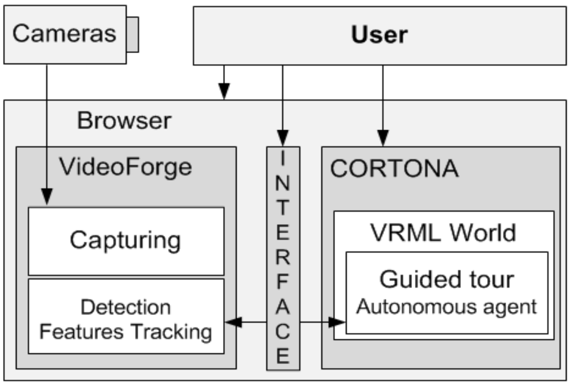

Figure 5. Simple block diagram of the system.

5. CONCLUSIONS

We are introducing our working results used in projects dealing with delivering information to users through virtual environments. Our work is still in progress and some work has to be done to fulfill our final aim to create fully working emphatic user interface. Here we presented only our partial working results, but those already used in practice.

We are planning for the future to extend this face to high precision face that will be compatible with MPEG-4 definitions and will contain all feature points defined in MPEG-4.

Now we already know that to successfully capture main feature points on user face we need to use cameras with higher resolution. This way we will need to have bigger processing power to work in real time. In the future this system should be placed and work on the Internet. We expect sending huge amount of data to users over Internet or establishing distant communication between users.

There is a group of people dealing with similar ideas how to extend user interface for viewing virtual environments with cultural heritage and they are also using capturing system. Their work is presented in [FRAN08]. Hopefully in close future we will exchange useful information and we will try to cooperate with this group on a project dealing with user interfaces for virtual environments with cultural heritage.

6. ACKNOWLEDGMENTS

The authors gratefully acknowledge the Scientific Grant Agency VEGA for supporting this work under the contract "Complexity of Geometric Algorithms for Real time rendering in Virtual reality" No. VEGA 1/3083/06.

7. REFERENCES

[ABR99] ABRANTES, G. - PEREIRA, F. 1999. MPEG-4 Facial Animation Technology: Survey, Implementation and Results. IEEE CSVT vol. 9, no. 2, pp. 290-305, 1999.

[HAnim] H-ANIM. Humanoid ANIMATION WORKING Group. http://www.h-anim.org/ [online] Accessed October 23, 2008.

[FER04] Ferko, A. et AL.. 2004. Virtual Heart of Central Europe. CORP 2004. www.corp.at. Vienna: TU Wien 2004. [online] Accessed October 23, 2008.

[GEJ04] GEJGUŠ, P. – KUBÍNI, P. 2004. Face tracking using convolution filters and skincolor model. In Proceedings of International Conference on Computer Vision and Graphics. Berlin:Springer, 2006. ISBN 1-4020-4178-0. pp. 721-726. Warsaw, 22.-24.9.2004. PL

[KUB06] KUBÍNI, P. 2006. VideoForge - An image processing system. In Proceedings of Spring Conference on Computer Graphics 2006: Conference Materials Posters. Bratislava: Comenius University, 2006. pp. 69-70 SCCG Častá-Papiernička , 20.-22.4.2006.

[PAR96] PARKE, F. I. – WATERS, K. 1996. Computer Facial Animation, ISBN 1-56881-014-8, A K Peters. Ltd. 1996.

[PER00] PERLIN, K. 2003. Face demo applet using noise. SIGGRAPH 2000. [online] http://mrl.nyu.edu/~perlin/facedemo. Accessed October 23, 2008.

[PMZ08] PAGE OF Považské MÚZEUM ŽILINA. 2008. www.pmza.sk [online] Accessed October 23, 2008.

[QVO02] QVORTRUP, L. ed. 2002. Virtual Interaction: Interaction in Virtual Inhabited 3D Worlds. London Berlin Heidelberg: Springer-Verlag 2002. ISBN 1-85233-516-5.

[QVO01] QVORTRUP, L. ed. 2001. Virtual Space: Spatiality in Virtual Inhabited 3D Worlds. London Berlin Heidelberg: Springer-Verlag 2001. ISBN 1-85233-331-6.

[STA02] STANEK, S., FERKO, A. 2002. Navigation and Interaction in Cyber Cities: Head Motions and

Facial Expressions. Pp.75-78 in Proceedings of Symposium on Computational Geometry SCG 2002. Slovak University of Technology in Bratislava 2002. ISBN 80-227-1773-8

[STA03] STANEK, S. - FERKO A. - KUBINI, P. 2003. Real-time Virtual Storytelling for Augmented Cultural Heritage: Message & Empathy. Pp. 45-46 in MAGNENAT-THALMANN, N. ed. Proceedings for the first Research Workshop on Augmented Virtual Reality. Organized by MIRALab, University of Geneva, September 18th-19th 2003, Geneva

[VHCE04] International EC Culture 2000 project: Virtual Heart of Central Europe. www.vhce.info. [online] Accessed October 23, 2008.

[VRML97] The Virtual Reality Modeling Language, VRML97. Functional specification and VRML97 External Authoring Interface (EAI) International Standard. ISO/IEC 14772-1:1997 and ISO/IEC 14772-2:2002.

[ŽÁRA02] ŽÁRA, J. 2002. Concise Tour to the Virtual Old Prague.,. EUROGRAPHICS 2002, Short Presentations, Saarbrucken:. Eurographics Association, 2002, pp. 191-198.

[FRAN08] FRANGESKIDES, F., LANITIS, A. and PAPANTONIOU, G. 2008.A contact-less interactive tool for exploring archaeological data. pp. 307-310 in Digital Heritage - Proceedings of the 14th International Conference on Virtual Systems and Multimedia. October 20th-25th 2008, Limassol, Cyprus