

## Virtual Lady Ada, an interactive museum exhibition

Simon Senecal, Nadia Magnenat Thalmann

### *1 Introduction*

The restitution of an intangible scene is basically composed in two different tasks: Reconstructing the tangible heritage of the cultural site and proposing a restitution of the intangible cultural heritage. See [1] and [2] for a survey of the possible way to present a general classification of the different approaches that might be employed to constitute a visual representation of cultural heritage. A lot of previous work has already been done in the domain of virtual tangible heritage reconstruction, see [3], [4] and [5]. In this specific area, different points are critical: the digitization and processing of the environment, the preparation of the lighting model and the modelling of the textures. In order to reach an immersive cultural VR or AR application including intangible content, the critical aspects are the modelling and animation of the virtual character. Indeed, the restitution of scenes involves virtual characters [6] and [7].



*Fig. 1: 17<sup>th</sup> century painting and 3D reconstruction of Lady Ada Byron*

The last 2 decades, a lot of work has been done in the domain of virtual museum. Tangible and intangible restitution in virtual museums has to be historically precise, pedagogic and well designed. The additional difficulty is to provide an efficient application that satisfy the user's expectations. In this chapter we present the different method around interactive behavior for virtual humans. We illustrate technics based on our work for the special Lady Ada Byron's exhibition at the Nixdorf Museum in Germany (see figure 22 and [8]). A virtually modelled Lady Ada was projected onto a big wall screen inside a tunnel. She could react to the presence of people and talk to them, explaining historical clues related to her story.

We will present first the tracking development that allow to feel the visitor, and then how this information is integrated to a global behavior system that give constituency to the virtual human and embed all necessary components.

## ***2 Tracking***

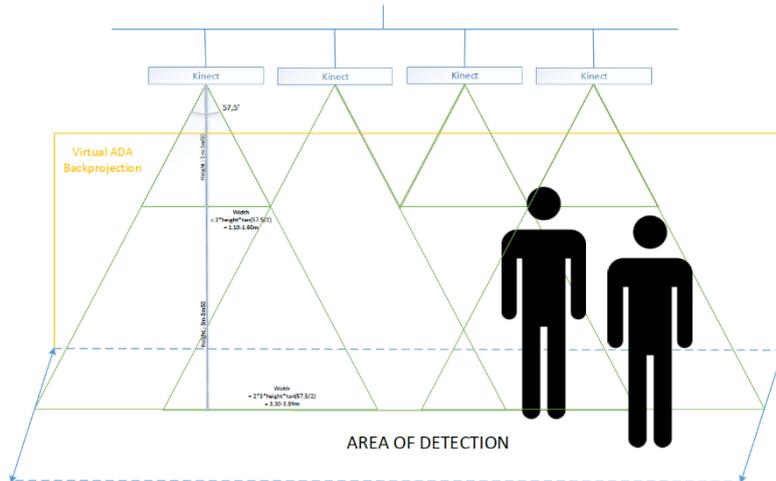
To give senses to our VH we need to get a stream of information about the behavior of the passing people. Usually this means the use of sensors that can extract various information from the visitors, along the time and with a sufficient accuracy. In our case study we needed the position of the people at least 15fps. This was made possible by the use of multiple sensors and tracking technology. The device Kinect V1 [9] (Microsoft Xbox One), which was originally developed for home entertainment purposes, but has also gained popularity in the scientific community [10 and 11]. The Kinect is an Infrared depth scanner and HD camera (720p), composed of an infrared (IR) projector, an infrared camera and an RGB camera.



*Fig. 2: Tracked Hand. The position of the center of the hand is determined in real time by a Microsoft Kinect V1 that provide depth information*

## **1 Acquisition**

We set up 4 Kinect for windows v1 side by side, such that the area covered by each Kinect is along the other one. All connected to a computer that will process the information and output on a video projector.



*Fig. 3: Schematics of the sensors set-up. On top are the four Kinects that detect people in a contiguous circular area.*

## 2 Processing

The incoming data acquired is four images of depth at a variable rate. We then have a buffer that wait to receive four images to then merge them into one to avoid missing images. We extract blobs based on the depth information. We decided to use a depth corresponding to a range of 40cm, between 1.50 and 1.90m above the ground to be able to detect most people. An additional function give us the position of the center of this detected blob over time. Another memory function insure we know the position of the same person over time without misleading the person. The output is an array of person position over time, so their trajectory.

## 3 Analysis

From the position of the user at a certain rate, we then processed these information to extract high level information. Using Markov chains and statistical analysis, we could define a spatial behavior of people corresponding to their intentions based on a sequence of position (around 40 positions). Therefore we know if the people are approaching the virtual human, going away or staying at their place. Two different information are outputted to the behavior system:

- The spatial region where the people are at any time (near the screen, on the right, etc..)
- The intention of the people that come from an analysis of the sequence of motion

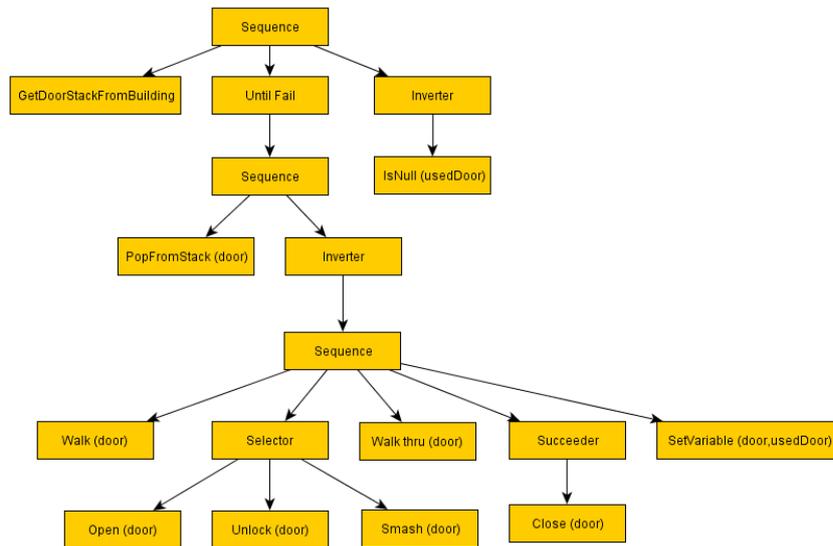
These two information are produced at different frequency, and will be integrated in the virtual human behavior by replacing values when these values are changing. Therefore the behavior loop is time-independent of the tracking unit.

#### ***4 Behavior Integration***

The locomotion, gestures, speech and reaction of the avatar is driven by a behavior management unit that guarantee the constituency and synchronization of the virtual human visual behavior. In the case of human-machine interaction for entertainment purposes, the people are expecting a fast and reliable feedback of the interaction device, which is in this case the virtual human, as suggested in [12], [13] and [14].

##### **1 Behavior Management system**

The synchronization between the different possible actions of the virtual human and a huge reliability is essential for his credibility. Such objectives can be achieved using behavior trees. Behavior tree is a programming technics that allows management of errors and is well suited for real time application.



*Figure 3: Example of a behavior tree. Here it illustrate how to open a door.*

It is made of hierarchical states that can be an action, animation or modification of variable. The behavior made in our case is oriented interaction: The VH passes from idle states where he is doing nothing or static actions to interaction mode where he is approaching the screen as the people are coming and wave at them before talking.

## 2 Gaze

Having the virtual human looking at you even if you are moving is a common features that is very important for credibility. We use the position tracking and projected that into a plane so that every person in front of virtual Ada will feel that she is looking toward them (figure 26). An adjustment in depth is also necessary to avoid looking above people or in case of children.



*Figure 4: Virtual Ada face close-up. The eyes follow the spectator motion.*

### 3 Locomotion

The motion of the virtual Ada should be related to the people behavior. Indeed, having the VH walking toward you when you approach him make sense and is part of the believability. However, it is not necessary that virtual Ada get close to the people. It can also run away from them or other behavior, as long as we feel the VH is somehow aware of the visitors. We uses navigation system with pathfinding and root motion. Different technics can be taken for moving in space:

- Using the spatial motion embedded in the animation, if it is a long clip. It means we don't have precise control every second but roughly every 10s.
- Using Pathfinding and walk animation. This means to have already walk animations. Then we can choose dynamically a place to go and the VH will move accordingly. Here again, there is two ways of doing:
  - First move the virtual human in translation and/or rotation in the same time as the animation is playing. These two has to be synchronized to avoid sliding effect.
  - Using so called root motion. This allows the translation to be determined by the animation, which make more sense and much more easy to use, because no tuning is required. However in that case the driving of the animation by the pathfinding side should be well thought.

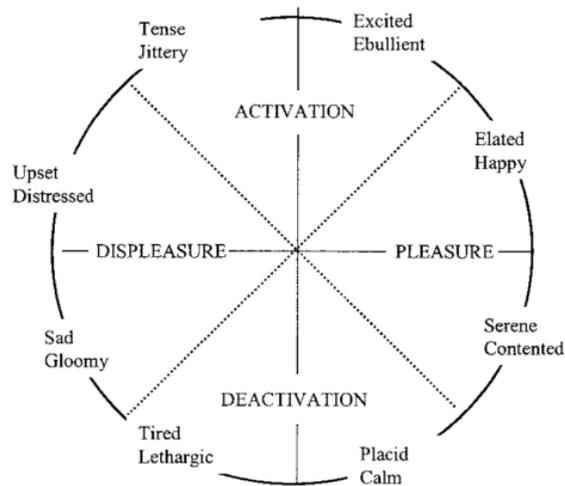


*Figure 5: Virtual Ada walking in determined direction driven by the pathfinding algorithm*

For virtual Ada, we defined six main spots that she is moving toward. The locomotion behavior consist then in moving to front spots when people are present, and going to back spot when nobody is there (figure 27). The gaze motion take control of the joints around the head and overlay other animation such as lip syncing for talks, allowing constituent multiple animation at a time (for example talking and moving the head toward someone).

### 5 Emotion recognition

Using the studies from [15], we integrate an emotion recognition system based on body movement analysis. This allows the avatar to be aware of the emotion expressed by the people trying to interact with him. It will be then used during the avatar’s talk to better communicate with the user.

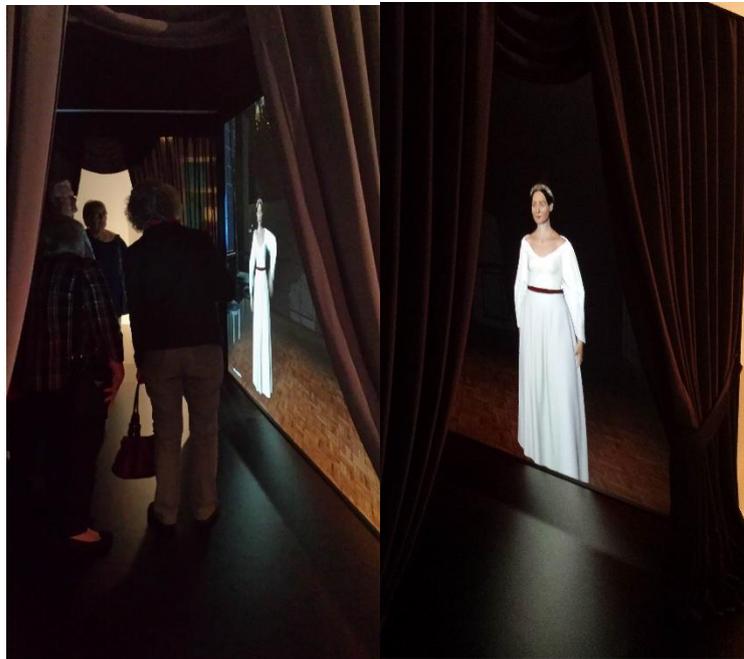


*Figure 6: Russel continuous model of emotion. Each emotion is defined by a two dimension coordinates, Pleasure and Arousal.*

The emotion recognition system is based on an emotion model from one hand and a motion model on the other hand. A complete motion analysis is made in real time then processed using statistical analysis. The output is an estimation of each people’s emotion regarding the Russel’s classification. The motion of the virtual Ada will be influenced by the recognized emotion.

## ***6 User feedback and application***

Lots of adjustment has to be done on-site as the timing of the virtual human behavior needs to be tuned when testing in real condition with behavior of multiple people. The real time analysis and synthesis can be tricky and could go fastly to some bugs or blocked states. This is why the program has to be bug-resistant and running even if some part of it are stuck. The result of the Ada exhibition (figure 29) is an interactive virtual human that react in function of the visitor motion inside a specific area. Position tracking, emotion recognition and behavior analysis is been used to determine a precise state of the people.



*Figure 7: Virtual character interaction with passing people in a tunnel. Sensor are located above the people and allow the interaction. Virtual Ada thanks the people to have been to the exhibition and wishes good way back.*

## **5 CONCLUSION**

Bringing life in a cultural context means to take into account a lot of information and constraints. Modelling a full human with his accessories and his clothes is already a challenge. We need then to model the environment and integrate the Human inside the environment. Then the motion correlated with gesture and interaction has to be brought. Many decision has to be made at different level that

have consequences. The choice depend on what is the objective: what we want to show and what is the material at disposition. Depending on the way we want to interact (Kinect, position sensor, gesture sensor, distance, speech recognition) we can organize a different virtual human behavior.

We showed through our examples corresponding to multiples European and side project that each case study will drive the development of the final application or research. For cultural heritage, the different resources to digitize can be complex and needs careful attention. Furthermore, the interactivity should follow the educational objectives as for a specific use [16]. The new technics and devices allows now more virtual and augmented reality tours that can enhance the user's experience.

#### REFERENCES

- [1] Marlène Arévalo, Nedjma Cadi-Yazli, Nadia Magnenat Thalmann. Progress in Cultural Heritage. Documentation, Preservation, and Protection 5th International Conference, EuroMed 2014, Limassol, Cyprus, November 3-8, 2014, Proceedings.
- [2] Alessandro E. Foni, George Papagiannakis, N. M.-T. (2010). A taxonomy of visualization strategies for cultural heritage applications. *Journal on Computing and Cultural Heritage (2010)*, 3(1), 1–21
- [3] Foni, A. E., Papagiannakis, G., Cadi-Yazli, N., & Magnenat-Thalmann, N. (2007). Time-Dependent Illumination and Animation of Virtual Hagia-Sophia. *International Journal of Architectural Computing*, 5(2), 283-301.
- [4] Magnenat-Thalmann, N., Foni, A. E., Papagiannakis, G., & Cadi-Yazli, N. (2007). Real Time Animation and Illumination in Ancient Roman Sites. *IJVR*, 6(1), 11-24.
- [5] Papagiannakis, G., & Magnenat-Thalmann, N. (2007). Mobile augmented heritage: Enabling human life in ancient pompeii. *International Journal of Architectural Computing*, 5(2), 396-415.
- [6] Egges, A., Papagiannakis, G., & Magnenat-Thalmann, N. (2007). Presence and interaction in mixed reality environments. *The Visual Computer*, 23(5), 317-333.
- [7] Apostolakis, K. C., Alexiadis, D. S., Daras, P., Monaghan, D., O'Connor, N. E., Prestele, B. & Moussa, M. B. (2013, July). Blending real with virtual in 3DLife. In 2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) (pp. 1-4). IEEE.
- [8] <http://www.hnf.de/en/sonderausstellungen/preview-ada-lovelace.html>
- [9] <https://developer.microsoft.com/en-us/windows/kinect/develop>
- [10] Kitsikidis, A., Dimitropoulos, K., Douka, S., & Grammalidis, N. (2014). Dance Analysis using Multiple Kinect Sensors. *VISAPP2014, Lisbon, Portugal*, 789–795.
- [11] Mao, Q.-R., Pan, X.-Y., Zhan, Y.-Z., & Shen, X.-J. (2015). Using Kinect for real-time emotion recognition via facial expressions \*. *Front Inform Technol Electron Eng*, 16(4), 272–282.
- [12] Moussa, M. B., & Magnenat-Thalmann, N. (2013). Toward socially responsible agents: integrating attachment and learning in emotional decision-making. *Computer Animation and Virtual Worlds*, 24(3-4), 327-334.
- [13] Gutiérrez, M., García-Rojas, A., Thalmann, D., Vexo, F., Moccozet, L., Magnenat-Thalmann, N. & Spagnuolo, M. (2007). An ontology of virtual humans. *The Visual Computer*, 23(3), 207-218.
- [14] Kasap, Z., & Magnenat-Thalmann, N. (2007). Intelligent virtual humans with autonomy and personality: State-of-the-art. *Intelligent Decision Technologies*, 1(1, 2), 3-15.

- [15] Simon, S., Cuel, L., Aristidou, A., & Magnenat-Thalmann, N. (2016). Continuous body emotion recognition system during theater performances. *Conference of Animation and Social Agents, CASA*.
- [16] O'Connor, N. E., Tisserand, Y., Chatzitofis, A., Destelle, F., Goenetxea, J., Unzueta, L. & Thalmann, N. M. (2014, September). Interactive games for preservation and promotion of sporting movements. In 2014 22nd European Signal Processing Conference (EUSIPCO) (pp. 351-355). IEEE.