Performance-Driven Facial Animation with HALCA and XVR

FEUP
Faculdade de Engenharia

Ricardo Marin (pdd10013@fe.up.pt)
Digital Medias, UP

FCT
Fundação para a ciência e a Tecnologia

Introduction

Being the dominant mean of human communication, facial expressions and motion are regarded as one of the fundamental topics in computer animation. Applications are among interactive games, using live real motion to control a character's face; virtual reality and animation. In a performance-driven animation system, usually the workflow is separated in three main components: Motion data acquisition with *Motion Capture* (MoCap), design of a *Facial Rig* to control the animation, and *Mapping/Retargeting* to the defined rig.

Towards the goal of finding a standard method to relate facial motion data to rig parameters, in order to be able to perform retargeting to different kinds of faces' topologies, we have currently implemented a marker-based first approach to synthesize facial expressions in real time. By using the Optitrack ARENA software [2], the animation library HALCA [3], and the XVR development framework [4], we are able to synthesize in real time local facial deformations on an a priori modeled and rigged 3D face model.

Methods

Our system is composed of 3 different components: The ARENA expression, a Client Application and an XVR Project. The ARENA software allowed us to administrate facial motion capture using facial markers. We used a previously fully calibrated standard set of six Optitrack cameras to define our volume of capture; and 24 facial markers (including head tracking) to define a facial template to track, which depended upon the definition of the facial rig for the 3D model to animate.

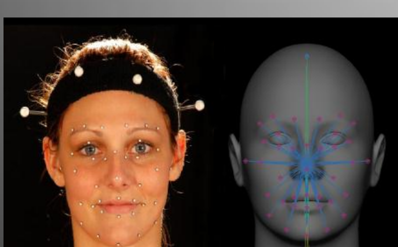

Figure 1. *Markers distribution example and its corresponding rig based on hierarchical joints*

For performance-driven animation in real-time, we used the streaming capabilities of ARENA in conjunction with its C++ NatNet SDK by implementing a compatible client application. This last was developed to retrieve in real time the facial template information created previously in ARENA, along with 3D markers positions for each of the frames, scaling of the data, correct orientation and head tracking. This data is finally sent via UDP to an XVR application, which in turn implemented the HALCA library to load all animations functions and controls over the avatar mesh and rig.

In our procedures we used a joint based facial rig and a direct mapping between joints and facial markers (Figure 1).

Mapping Process

Let $R = \{J^t_i\}$ be the facial rig of our model, defined as a set of n facial joint positions at time t. Let $M = \{M^t_i\}$ be the set of corresponding n markers positions also at time t, without counting the four head tracking markers. Suppose that at $t=0$, J^0_i stores the neutral expression joint positions, and similarly M^0_i stores the neutral user expression markers positions. In a direct mapping procedure, it is natural to relate displacements of the joints to displacement of the markers. Therefore a linear model can be constructed by the equation

$$J^t_i = J^0_i + K_i(M^0_i - M^t_i)\ ,\ i=1,..,n;\ \text{for all } t$$

Which allow us to get joints positions for all time (frames) t. The constant factors K_i could be modeled to influence more certain parts of the face than others that we know are more stable along time. The range of motion calibration can be attached to the 3D model rig definition procedure, simulated with predefined values for these factors by regions. This mapping method however could not be applied to joints whose motion is due to rotation rather than translation, e.g., the Jaw joint. For this, we used a cosine law to solve for rotation angles in joints from displacements in markers, in a very similar fashion to the aforementioned linear method. The head rigid motion is extracted from four headband markers from Optitrack.

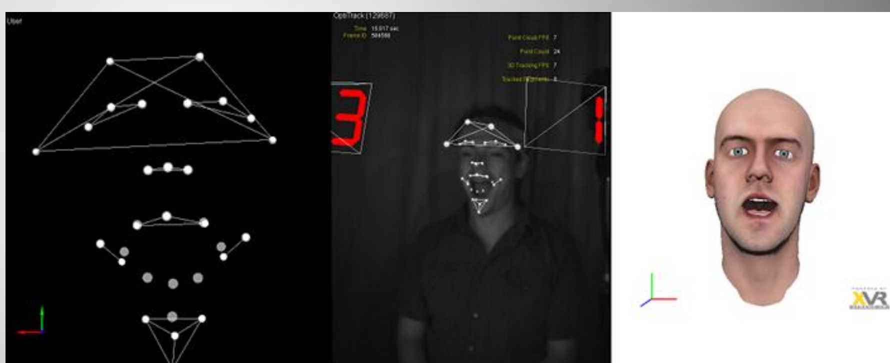

Figure 2. *Performance-driven Animation: Facial Template(Left), marker based mocap (Middle), and animation (Right)*

Results - Work in Progress

We performed two MoCap sessions to test our mapping and animations in real-time. In general the range of motion factors K_i needed to be adapted for each of the participants, in a range value from 0.25 to 1. The assumption of linearity between joints and markers displacements synthesized facial expressions in real time accordingly to live performance but losing subtleness of a particular actor. Artifacts however were caused mainly by markers tracking instability, and bad weighting of the joints.

In order to get the subtleness within the expressions of each user, a different mapping needs to be developed, including the design of another rig. Current development points to address: Optimal number of markers that are needed to capture subtleness of facial expressions. Optimal facial rig and skinning accordingly to the number of markers: Shape-Aware joint weighting as an alternative to the traditional manual painting; and finally to replace reflective markers for feature tracking in a vision based method such as Active Appearance Models (AAM).

Acknowledgements

This research is supported by the European Union FP7 Integrated Project VERE (No. 257695) Instituto de Telecomunicaçoes - IT and Fundação para a Ciência e a Tecnologia - FCT, Portugal.

References

[1] Alexander, O., M. Rogers, et al. (2009). *The Digital Emily project: photoreal facial modeling and animation.* ACM SIGGRAPH 2009. New Orleans, Louisiana, ACM: 1-15.
[2]http://www.naturalpoint.com/optitrack
[3]http://www.lsi.upc.edu/~bspanlang/animation/avatarslib/doc/
[4]http://wiki.vrmedia.it/index.php?title=Main_Page

Virtual Embodiment and Robotic Re-Embodiment
instituto de telecomunicações
UT Austin | Portugal
International Collaboratory for Emerging Technologies - CoLab