Performance-Driven Facial Animation with HALCA and XVR

Ricardo D. C. Marín
Facultade de Engenharia – IT/FEUP
Rua Dr. Roberto Frias, 4200-465
Porto, Portugal
+351 225 081 400
pdd10013@fe.up.pt

Xenxo Álvarez
Instituto de Telecomunicaçoes - IT
Rua de Campo Alegre 1021/1055
Porto, Portugal
+351 220 402 931
xenxo3d@gmail.com

Veronica Orvalho
Instituto de Telecomunicaçoes – IT/FCUP
Rua de Campo Alegre 1021/1055
Porto, Portugal
+351 220 402 931
coveronica.orvalho@dcc.fc.up.pt

Introduction

Facial expressions are the dominant mean of human communication and characterization, and thus their computational simulation and animation on 3D avatar faces has become the driving force of any research confronting the problem of facial animation. In particular within the context of the VERE project, a real-time and marker-less performance-driven facial animation must be achieved in order to make a closest-to-real immersion experience, for simulating social interactions where the user's facial expressions are synthesized on his virtual (face) representation in a synchronized fashion.

In these performance-driven animation systems, 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. State of the Art methods based on real time photometric stereo and structured light scanners can integrate all three procedures capturing geometry, textures and motion at once [1], [2]; they are however not suited for our purposes since they have several manual processing interventions, including the retargeting to different kinds of faces' topologies, which is an additional feature needed for experimentation and research within VERE. In this case, mapping motion capture data to a rig becomes as fundamental as the recording technology, and a standard method to relate facial motion data to rig parameters needs to be addressed.

Towards this goal, in this paper we present a marker-based first approach to synthesize facial expressions in real time. By using the Optitrack ARENA software [3], the animation library HALCA [4], and the XVR development framework [5], 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. 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.

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

$$J^t_i = J^0_i + K_i(M^0_i - M^t_i)\ ,\ i=1,..,n$$

which allow us to get joints positions for all times 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. This 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.

Results

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 satisfactorily synthesized facial expressions in real time. Artifacts however were caused mainly by markers and tracking instability.

Conclusions and Future Work

In general our synthesized facial expressions resembled that of the participant's including particular facial deformations. However, in order to get the subtleness inside the expressions of each user, a different mapping needs to be developed, including the design of another rig (e.g. blendshapes). One of the key points of our method is that it can be integrated with computer vision approaches to facial tracking (Active Appearance Models-AAM) without major changes, given that these kinds of methods uses virtual markers (features) that can be mapped to the avatar's rig.

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] Fyffe, G., T. Hawkins, et al. (2011). "Comprehensive Facial Performance Capture." Computer Graphics Forum **30**(2): 425-434.

[2] 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.

[3] http://www.naturalpoint.com/optitrack

[4] http://www.lsi.upc.edu/~bspanlang/animation/avatarslib/doc/

[5] http://wiki.vrmedia.it/index.php?title=Main_Page