GALA 2007 submission document

Title: Avatar puppetry using real-time audio and video analysis

Track: Student
Category: Animated Lifelike Agent Application
Authors:

Sylvain LE GALLOU
3rd year PhD Student
Rennes University, France

Gaspard BRETON
R&D Engineer,
France Telecom R&D, France

Renaud SEGUIER
Associate Professor
Supélec SCEE/IETR, France

Christophe GARCIA
R&D Engineer,
France Telecom R&D, France

Contact data:
name: Sylvain LE GALLOU
e-mail: sylvain.legallou@orange-ftgroup.com
url:
phone: +33 299 124 468
postal address: *4 rue du Clos Courtel*
BP 91226,
35512 Cesson-Sévigné, France

URL:

Movie file submitted: *LE_GALLOU.AVI*

Reference teacher:

Gaspard BRETON**.**
affiliation: France Telecom R&D/TECH/IRIS/IAM
e-mail: Gaspard.Breton@orange-ftgroup

Process of development: The Avatar puppetry system was developed in March 2007, as a project to highlight the capabilities of the head tracking algorithm developed by the first author (Sylvain Le Gallou) in collaboration with the 3 other authors (Gaspard Breton, Renaud Séguier and Christophe Garcia).
The 2nd author (Gaspard Breton) was contacted to help with the animation system and the body design.

Resources used:

1. Matlab for face deformable model creation used by video analysis
   http://www.mathworks.fr/
2. VisualC++ 6, OpenCv and OpenGl for programming

Resources required: The system was developed and tested on a Laptop with Windows XP Prof OS, 1 Gb RAM, NVIDIA Quadro NVS graphics card. The PC's microphone and a Logitech webcam attached were used.

GALA 2007 submission document

Avatar puppetry using real-time audio and video analysis

1. The application and context of the work

We present a system which consists of a lifelike agent animated in real-time using video and audio analysis from the user. This kind of system could be used for Instant Messaging where an avatar controlled like a puppet is displayed instead of the webcam flow.

2. Novelty

In order to do video analysis of faces, we carried out a face alignment based on the Active Appearance Model (AAM) method [1]. AAM is a deformable model method which allows shape and texture to be jointly synthesized by statistical shape and texture models.

We also implemented a preprocessing step in order to improve the robustness of the AAM illumination variations. Indeed we carried out the adaptive histogram equalization as a preprocessing on images before the AAM method like [2]. This method provides a good video flow analysis of faces without limitations in background and illuminations.

The audio analysis is performed using a common HMM model learned from several speakers in a noisy environment. The system is multi-speaker and works for both genders. The phonetic segmentation is performed in real time on 64ms buffers allowing a really short delay.

The animation is performed using the FaceEngine 3D animation system [3] which works in real-time. The head movements are computed using a behavior engine [4] taking into account biological constraints such as the vestibulo-ocular reflex and the head inertia. Lips movements are performed using a co-articulation algorithm blending the visemes corresponding to the phonetic string returned by the audio analysis.

3. The architecture

The overall system is made of video analysis based on Active Appearance Models and audio analysis based on Hidden Markov Model. The parameters from these two modules are sent to a control system driving the animation engine. The video analysis extracts the head orientation and the audio analysis provides the phonetic string used to move the lips.

4. Performance

Audio and video analysis and the animation system are performed in real-time without any perceptible delay.

References

[1] T.F. Cootes, G.J. Edwards, C.J. Taylor: "Active Appearance Models", ECCV'98, European Conference on Computer Vision (1998).

[2] S. Le Gallou, G. Breton, C. Garcia, R. Séguier: "Distance Maps: a robust illumination preprocessing for active appearance models", VISAPP'06, International Conference on Computer Vision Theory and Applications (2006).

[3] G. Breton, C. Bouville, D. Pelé, "FaceEngine: "A 3D Facial Animation Engine for Real Time Applications", presented at Web3D Symposium, Paderborn, Germany (2000).

[4] G. Breton, D. Pelé, C. Garcia: "Modeling gaze behavior for a 3D ECA in a dialogue situation", Proceedings of the 11th international conference on intelligent user interfaces, Sydney, Australia (2006).