Avatar Puppetry Using Real-Time Audio and Video Analysis

Sylvain Le Gallou[1,2], Gaspard Breton[1], Renaud Séguier[2], and Christophe Garcia[1]

[1] France Telecom, TECH/IRIS Team,
rue du Clos Courtel, Cesson-Sévigné, France
firstname.lastname@orange-ftgroup.com
[2] Supélec/IETR, SCEE Team, avenue de la Boulaie,
Cesson-Sévigné, France
firstname.lastname@supelec.fr

Abstract. 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. 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.

1 System Overview

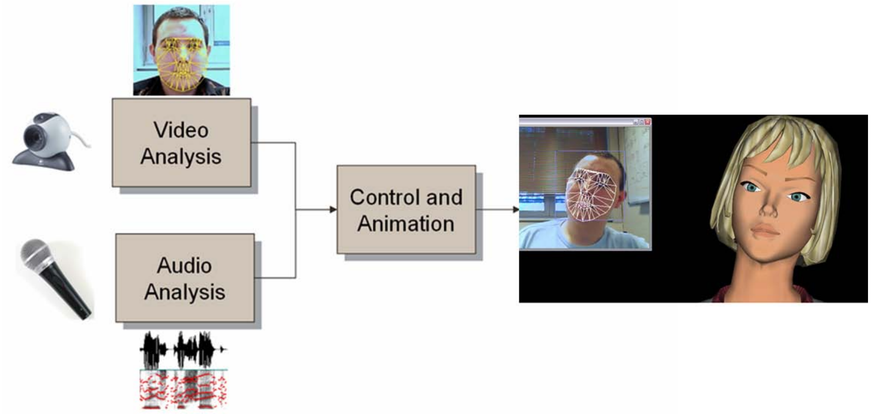

Fig. 1. Overview of the pupettry system

C. Pelachaud et al. (Eds.): IVA 2007, LNAI 4722, pp. 391–392, 2007.

2 Video Analysis

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. The creation of the statistical shape and texture model is performed by learning from a database of different examples of faces. In order to be able to compute the orientation of the head, the model has learned from a database containing various faces with different expressions and orientations. 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.

3 Audio Analysis

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.

4 Control and Animation

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 [5] blending the visemes corresponding to the phonetic string returned by the audio analysis.

References

1. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, Springer, Heidelberg (1998)
2. Le Gallou, S., Breton, G., Garcia, C., Sguier, R.: Distance Maps: a robust illumination preprocessing for active appearance models. VISAPP'06, International Conference on Computer Vision Theory and Applications (2006)
3. Breton, G., Bouville, C., Pel, D.: FaceEngine: A 3D Facial Animation Engine for Real Time Applications. Web3D Symposium, Paderborn. Germany (2000)
4. Breton, G., Pel, D., Garcia, C.: Modeling gaze behavior for a 3D ECA in a dialogue situation. In: Proceedings of the 11th international conference on intelligent user interfaces, Sydney, Australia (2006)
5. Cohen, M.M., Massaro, D.W.: Modeling Coarticulation in Synthetic Visual Speech. Models and Techniques in Computer Animation. Springer, Heidelberg (1993)