Online Face Analysis: Coupling Head Pose-Tracking with Face Expression Analysis

Ana C. Andrés del Valle Jean-Luc Dugelay
Institut Eurecom
Multimedia Communications Department
2229 route des Crêtes. 06904 Sophia Antipolis. France
+33 (0)4.93.00.{29.21}{26.41}
{ana.andres}{jean-luc.dugelay}@eurecom.fr

ABSTRACT
Future human-machine interaction interfaces will need a perfect understanding of the person's behavior so that machines can learn from it, react accordingly, synthetically reproduce this behavior afterwards, etc. The study of user's head action and face expression is fundamental to achieve this comprehension. This technical demo shows the latest image processing techniques for face tracking and expression analysis developed at Eurecom. Our research on face analysis aims at synthetically reproducing head movements in telecom applications. Users will be able to test themselves on how our system can track and analyze their face expressions with just one webcamera only under any unconstrained environment.

Categories and Subject Descriptors
H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems – *Animations, Artificial, augmented, and virtual realities*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis – *Tracking, Shape, Motion*.

General Terms
Design, Experimentation, Human Factors.

Keywords
Face Analysis, Face Animation, Face Tracking, 3D-synthesis, Virtual Reality.

1. INTRODUCTION
Face video analysis is often performed through the study of specific face features (eyes, eyebrows and mouth) to extract the most significant information regarding expression and speech. Many of the current feature analysis algorithms are developed to work in 'near-to-front' face position. Whether the algorithms are used to do only expression analysis [1, 2] or they are oriented to perform Model Based coding [3], these systems do not allow the user to move freely.

Assuming that we control the user pose is an important restriction when doing analysis for videoconferencing purposes. Yet, most virtual telecommunication schemes [4,5] try to avoid the pose-expression coupling issue by minimizing its effects. Nevertheless, for their analysis algorithms to remain robust, they only allow the user to do slight movements.

Developing a video analysis framework where head pose tracking and face feature analysis are treated separately permits to design specialized image analysis algorithms adjusted to specific needs, feature characteristics, etc. For our work on virtual teleconferencing environments, we first developed a pose-tracking algorithm (based on Kalman filtering) that profits from a tight analysis-synthesis cooperation. We are able to **track and predict the pose** of the speaker frame by frame with the help of the synthesis of its realistic 3D head model (clone). In parallel, we design image analysis algorithms to **study the expression motion** from a head on a 'near-to-front' position, situation at which faces show most of their gesture information. We have also developed a pose-expression coupling method that enables us to utilize the expression analysis algorithms taking advantage of the pose prediction given by the Kalman filter. Our algorithms permit the user to easily test the method because:

- they perform almost in real-time;
- they do not need actual physical constraints, for instance, special equipment, face markers or specific lighting conditions; and
- they do not need any previous training stage to learn about the user's face characteristics.

This demo will simulate a virtual teleconference terminal. Users will test the pose expression coupling procedure. They will be recorded with a camera and then their eye and eyebrow movements will be synthesized on an avatar. This way users will experience how future teleconference systems could be.

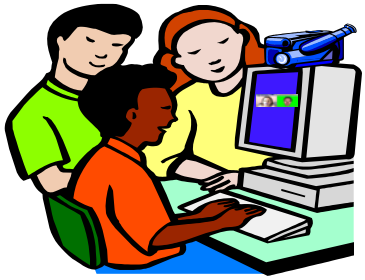

Figure 1. Settings for the demo: just one computer and one camera are needed.

Multimedia'02, December 1-6, 2002, Juan-les-Pins, France.

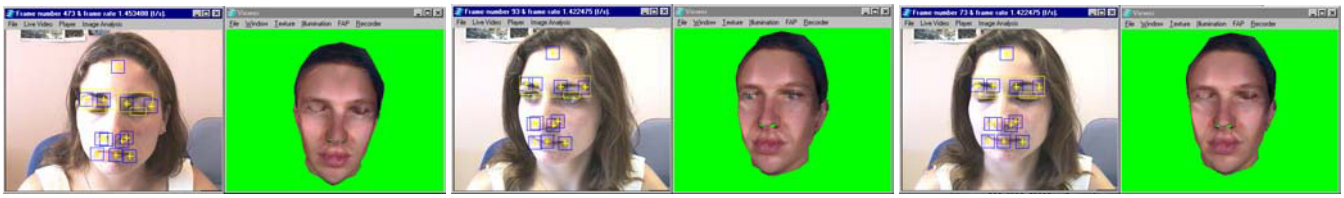

Figure 2. Feedback screens where users will be able to appreciate how the system analyzes their expressions and immediately reproduces them on the avatar.

2. POSE TRACKING: KALMAN FILTER

Kalman filters are often used in head tracking systems for two different purposes: the first one is to temporally smooth out the estimated head global parameters, the second one is to convert the positions of the 2D facial features observations into 3D estimates and predictions of the head position an orientation. In our application, the Kalman filter is the central node of our face-tracking system [6]: it recovers the head global position and orientation, it predicts the 2D positions of the features points for the matching algorithm, and –this is the point exploited for telecom applications– it makes the synthesized model have the same scale, position and orientation as the speaker's face in the real view, despite the acquisition by a non-calibrated camera.

3. FEATURE ANALYSIS ALGORITHMS

In our demo applications the analysis of eye and eyebrow expressions are coupled with the tracking of the head. Next, we briefly explain the image processing algorithms involved:

3.1 Eye State Analysis

We have developed a real-time lighting-independent eye movement estimation algorithm [7] that considers applying a combination of color and energy analysis algorithms using the HSI (Hue, Saturation, Intensity) pixel components of the feature image. Then we interpret the results of the analysis in terms of some specific action units that we associate to the temporal states. Following a logical state diagram we relate our analysis results to the final parameters that describe the eye movement.

3.2 Eyebrow Motion Analysis

To study eyebrow behavior from video sequences we utilize a new image analysis technique based on an anatomical-mathematical motion model. This technique [8] conceives the eyebrow as a single curved object (arch) that is subject to the deformation due to muscular interactions. The action model defines the simplified 2D (vertical and horizontal) displacements of the arch. Our video analysis algorithm recovers the needed data from the arch representation to deduce the parameters that deformed the proposed model.

4. R&D DEMO SETTINGS & DESCRIPTION

The tech-demo will have two major parts. First, users will be able to test our application online to see how eye movement is analyzed from the camera acquisition and then synthesized on an avatar. They will also be able to appreciate the eyebrow motion analysis (offline). For both parts, there will be just using a computer and a camera. Not special lighting, equipment or make-up is required.

Users will sit in front of a webcam situated on top of a monitor (simulating a typical Internet teleconference setting). During the demo users will have two feedback windows (see Fig. 2). On one of them, they will see the online video input (their face), on the other one, they will see the avatar synthesis. This demo starts with an initialization process where avatar and video input are rendered together. During this process users are supposed to fit their faces on that of the avatar.

5. REFERENCES

[1] F. Piat, and M. Tsapatsoulis, "Exploring the Time Course of Facial Expressions with a Fuzzy System," *ICME 2000*, Sydney - Australia, 27-30 Aug. 2000.

[2] M. Pantic, and L. J.M. Rothkrantz, "Automatic Analysis of facial Expressions: The State of the Art," *IEEE Trans. on Pattern Analysis and Machine Intelligenc*e, Vol. 22, No. 1 2, pp. 1 424-1 445, Dec. 2000.

[3] E. Cosatto, G. Potamianos, and H.P. Graf, "Audio-Visual Selection for the Synthesis of Photo-Realistic Talking-Heads," *ICME 200*0, Sydney - Australia, 27-30 Aug. 2000.

[4] T. Goto, S. Kshirsagar, and N. Magnenat-Thalmann, "Automatic Face Cloning and Animation," *IEEE Signal Processing Magazin*e, Vol. 1 8, No. 3, pp. 1 7-25, May 2001.

[5] P. Eisert, and B. Girod, "Analyzing Facial Expressions for Virtual Conferencing," *IEEE Computer Graphics & Application*s, pp. 70-78, Sep. 1 998.

[6] Stéphane Valente, and J.-L. Dugelay, "A Visual Analysis/Synthesis Feedback Loop for Accurate Face Tracking", *Signal Processing: Image Communication*, vol.16 (2001) pp. 585-608.

[7] A. C. Andrés, and J.-L. Dugelay, "Facial Expression Analysis Robust to 3D Head Pose Motion", *ICME 2002*, Lausanne – Switzerland, 26-29 Aug. 2002.

[8] A. C. Andrés, and J.-L. Dugelay,, "Eyebrow Movement Analysis over Real-time Video Sequences for Synthetic Representation", *submitted to AMDO 02*, Palma de Mallorca - Spain, 21-23 Nov. 2002