Efficient Ocular Expression Analysis for Synthetic Reproduction

Ana C. Andrés del Valle, *Member, IEEE,* and Jean-Luc Dugelay, *Senior Member, IEEE*

***Abstract*—This paper presents an original framework that analyzes the complete ocular (eye + eyebrow) expression on video sequences to reproduce it later on synthetic three-dimensional (3-D) models in real-time. We propose a two step process to develop robust techniques for facial feature analysis aimed at working without special illumination conditions or physical constraints on the user (markers, fixed frontal pose, etc.). First, simple and efficient image-processing methods based on motion models are designed over a frontal point of view of the face. Natural and realistic intra-feature and inter-feature constraints are applied to improve the analysis results. Then, the usability of these algorithms is extended to enable the analysis regardless of the person's pose in front of the camera. This is achieved by redefining the motion models involved over the speaker's highly realistic synthetic representation (clone), by using a suitable observation model, and by predicting the head pose in 3-D, frame by frame.**

***Index Terms*—Animation, communications systems, image motion analysis, image processing, Kalman filtering, machine vision, real-time systems, teleconferencing.**

I. Introduction

TWO premises have driven research in videoconferencing. Technically, more efficient ways of coding and transmitting video information are being sought ([1]–[3]); socially, new communication frameworks where visual information increases speaker interaction are under investigation [4]. New telecommunications trends consider achieving these goals by using synthetic data. For instance, in virtual teleconferencing, speakers are substituted by three-dimensional (3-D) synthetic models—clones (realistic) or avatars (symbolic). Classical video data is replaced by a limited number of action parameters determining facial motion, thus reducing the bandwidth required for transmission. Furthermore, natural communications are recreated when head models are synthesized in a common interactive environment.

The understanding of nonverbal facial behavior on monocular images, especially of eyes [5] and eyebrows, becomes critical to generate natural and coherent facial animation on synthetic head models when classical teleconferencing input (i.e., video) is used.

Manuscript received August 4, 2003; revised November 23, 2004. This work was supported in part by France Telecom R&D. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Richard B. Reilly.

A. C. Andrés del Valle was with the Eurecom Institute, BP 193, F06604 Sophia Antipolis, France. She is now with Accenture Technology Labs, BP 99, F06609 Sophia Antipolis, France (e-mail: ana.c.andresdelvalle@accenture.com).

J.-L. Dugelay is with the Multimedia Communications Department, Eurecom Institute, BP 193, F06604 Sophia Antipolis, France (e-mail: jld@eurecom.fr).

Digital Object Identifier 10.1109/TMM.2005.861290

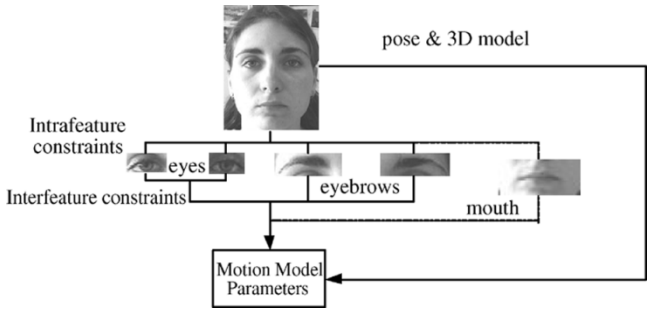

Fig. 1. General diagram of the proposed framework.

Many approaches that analyze feature behavior fail to provide a robust solution to all major requirements of common face-to-face communications: no control of lighting of the speaker's environment, complete freedom of movement in front of the camera, real-time capabilities, etc. The systems that permit the analysis and reproduction of head pose and facial expressions simultaneously use image analysis techniques (e.g., optical flow) that are highly dependent on the environmental conditions [1], [6]. Those that use more flexible and computationally less costly analysis methods [7], [8], usually developing specific motion analysis templates per feature, do not address the pose-expression interaction directly and try to minimize the effects of pose changes.

In this paper, we propose an eye and eyebrow motion analysis framework (Fig. 1) that meets the previously mentioned requirements for communications. The approach presented follows a two-step process to understand and generate coherent and natural eye and eyebrow motion.

— First, we design image-processing techniques to study the features extracted from a frontally-oriented face. The developed motion templates do not only utilize anatomical (intra-feature) constraints to derive the eye and eyebrow actions, as was done in [7]–[9]; they also use human natural standard motion constraints to generate natural realistic facial animation exploiting the existing correlation between eyes and between eyes and eyebrows (inter-feature constraints).

— Then, we extend our algorithms to analyze the features independently of the speaker's pose. In the literature, there are two solutions to adapt the use of motion templates when the speaker's pose is known.

  i) Either **one feature template is designed per pose**: after developing and testing motion templates on frontal faces, they are redefined based on different predetermined face poses [9]. This analysis strategy quickly becomes very limited. The complexity of the

solution increases with the number of states, which will be large if much freedom of movement is required.

ii) Or **the input image is rectified**: the image to be analyzed is transformed to obtain an approximation of the face viewed from a frontal perspective. Afterwards, the image processing algorithms defined for frontal faces analyze this new image to obtain the corresponding feature templates [10]. Strong rotations and translations cannot be compensated with such image transformations.

We have developed a novel solution for the adaptation of the studied eye and eyebrow motion templates. The knowledge of the head pose is used to interpret the expressions in 3-D space instead of on the image plane. To achieve this spatial understanding, the motion models and their parameters are redefined over the speaker's 3-D head model. The pose parameters are used to interpret the analyzed data by means of an observation model that describes the mathematical relationship between the coordinates of the head object in its 3-D neutral pose (facing the camera) and the final view of the face on the video in 2-D (image plane). This solution is the result of several years of research on realistic facial animation and telecommunications done at the Eurecom Institute.

This paper is organized as follows. The image processing motion templates are described in Section II and the inter- and intra-feature constraint considerations used to analyze eyes and eyebrows are detailed in Section III. Section IV explains how the proposed frontal pose analysis techniques are extended to still work for any other pose. Section V includes the experimental evaluation of the proposed framework. We conclude with remarks and future perspectives in Section VI.

II. Ocular Expression Analysis Algorithms for a Frontal View of the Face

Facial expressions are independent of head rigid motion. Although their projected appearance on the image is not completely independent of pose, the solution developed in this paper tries to exploit the real expression-pose independence in 3-D space to study ocular expressions in two-dimensional (2-D).

First, image-processing algorithms that study faces from a frontal perspective, in which they show most of their expression, are developed. These algorithms assume that location and delimitation of the feature region of interest (ROI) are known. This assumption is realistic in the present context because, as explained in Section IV, the procedure that extends the use of these algorithms to other poses also takes into consideration the tracking and definition of feature ROIs.

A. Eye State Analysis Algorithm

Natural ocular actions are characterized by the existence of a tight relationship between the vertical pupil location and the eyelid opening; therefore, we can expect to obtain much of the eye motion information from the analysis of pupil activity. The proposed analysis scheme reduces the study of eye motion to the determination of the pupil position within the eye area. The pupil is similar for all eyes, regardless of iris color. We assign an action state to the eye related to its position. Next, we review the image processing involved.

1) Pupil Search Algorithm: First, the eye region of interest is defined on the face, determining eye width ($Weye$) and pupil width ($Wpupil$). Then, we scan the ROI to find the point of minimum energy

$$
\begin{aligned}
&(X,Y)=(x,y)\\
&\quad s.t.\arg\min\left(\left[\frac{3}{4(\lambda\cdot Weye)^2}\right.\right.\\
&\qquad\left.\cdot\sum_{l=-\lambda\cdot\frac{Weye}{2}}^{\lambda\cdot\frac{Weye}{2}}\sum_{l=-\lambda\cdot\frac{Weye}{2}}^{\lambda\cdot\frac{Weye}{2}}I^2(x+l,y+m)\right]\\
&\qquad+\left[\frac{1}{4(\lambda\cdot Weye)^2}\right.\\
&\qquad\left.\left.\cdot\sum_{l=-\lambda\cdot Weye}^{\lambda\cdot Weye}\sum_{l=-\lambda\cdot Weye}^{\lambda\cdot Weye}I^2(x+l,y+m)\right]\right)\\
&\quad\forall x,y\exists\quad x+l>0,\quad y+m>0
\end{aligned}
\tag{1}
$$

where (X,Y) are the coordinates of the pupil location on the eye ROI and $I^2(x,y)$ is the energy of a pixel computed as the square of its intensity component. The evaluation formula tries to look for the intensity distribution that is closest to the pupil-iris shape on the analysis area. Since the appearance of the face on video may differ depending on the recording conditions, λ, the ratio $Wpupil/Weye$, is defined. Since λ is an anthropometric measure uniquely related to an individual, it remains constant for all scenarios and completely determines the evaluation.

2) Parameterization of the Analysis: To synthesize the results from the previous analysis technique, it is necessary to parameterize the obtained data. The parameterization process maps the (X,Y) pupil location onto their corresponding state values.

A state value S for a given frame at time t is obtained as a function of the location of the pupil with reference to the width W and the height H of the eye ROI that is being analyzed: $S^t = f(X^t,Y^t,W^t,H^t)$. To define the states, we divide the region of analysis into different zones and we assign to each of them a concrete eye action (e.g., look up, look down, look left, and so on). This action can be rendered if each of the states can be synthetically reproduced. The function assigns states following these constraints: a) we do not explicitly detect eyelid actions because eyelids partially cover the irises, and vertically follow the pupil motion; b) pupils always remain the darkest part of the eye when it is open regardless of the lighting conditions; and c) the absence of pupil indicates that the eye is closed, in such a case, the darkest point falls on the eyelashes. The assignment of states is a 2-D quantization process where the exact (X,Y) is mapped onto the closest pupil coordinate location for a given state.

A simplified example of this algorithm is proposed in [11].

B. Eyebrow Motion Analysis Algorithm

To study eyebrow behavior from video sequences, we utilize a new image analysis technique based on an anatomical-math-

Fig. 2. Several muscles generate the eyebrow movements. Upward motion is mainly due to the *Frontalis* muscle and downward motion is due to the *Corrugator*, the *Procerus* and the *Orbicularis Oculi* muscles. Images extracted from [12].

ematical motion model. This technique conceives the eyebrow as a single curved object that is subject to deformations due to muscular interactions. The action model defines the simplified 2-D 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.

1) Anatomical-Mathematical Eyebrow Motion Model: Although the shape of an eyebrow depends on the facial appearance of an individual, its motion is related to common muscular actions. This enables us to represent eyebrows as arches whose shape is specific to the person but whose motion is mathematically modeled following muscular constraints, i.e. speaker independent. The arch movement is parameterized as the displacement in the x and y-axis of each point that belongs to the arch compared to its initial neutral position, when no force acts: $\Delta x = x_t - x_{\text{neutral}}$, $\Delta y = y_t - y_{\text{neutral}}$.

Two different global behaviors exist in eyebrow motion, upwards and downwards. They are controlled by different muscles, so different motion models have been designed.

a) *Upwards*:

$$\Delta x = Ff_x \cdot e^{\left(-\frac{x_n}{\alpha}\right)} \tag{2}$$

$$\Delta y = Ff_y + Ff'_y \cdot e^{\left(-\frac{x_n}{\alpha}\right)}. \tag{3}$$

b) *Downwards*:

$$\Delta y = -Fcs_y + Foo_y \cdot \left(|x_n - \beta| - \beta\right)^2 \tag{4}$$

$$\text{If} \quad x_n < \beta \qquad \Delta x = -Fcs_x \tag{5}$$

$$\text{If} \quad x_n > \beta \qquad \Delta x = Foo_x \cdot (\beta - x_n) - Fcs_x. \tag{6}$$

Ff, Ff', Fcs, and Foo are the magnitudes associated to the force of the *Frontalis*, *Corrugator* and the *Orbicucularis Oculi* muscles, respectively, (Fig. 2). We refer to [12] for an anatomical description of the eye area. The action of the *Procerus* muscle, being highly correlated with the one from the *Corrugator*, is included in the Fcs term. Subscripts x and y indicate the different components; w represents the eyebrow width, $\alpha = 2w/3$ and $\beta = w/2$; and (x_n, y_n) are the eyebrow arch coordinates when no force acts. Fig. 3 presents the local eyebrow coordinate system. These expressions have been derived from the observation of the muscular motion of the eyebrows, the empirical study of the optical flow behavior of the eyebrow area observed on real video sequences and the adaptation of the parameters involved to the anatomical shape of the eyebrow. We preferred to develop our own model rather than using more complex models already available, because we wanted to control the influence of the image processing error on the study of the model behavior. More details about the algorithm development can be found in [13].

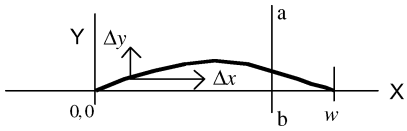

Fig. 3. Left eyebrow model arch and its coordinate reference. The origin for the analysis algorithm, it is always situated at the inner extreme of the eyebrow (close to the nose) and defined for the eyebrow in its neutral state.

$\Delta z = z_t - z_{neutral}$, the depth component of eyebrow motion, cannot be estimated from a frontal view and might be simulated by sliding the skin on the skull making the eyebrow follow the forehead surface. The animation technique "displacement of texture coordinate" described in [14] illustrates this concept.

2) Image Analysis Algorithm—Deducing Model Parameters: The developed image analysis algorithm tries to reduce the image of the eyebrow down to the proposed model in order to study the distribution of the points on the eyebrow arch. Then, it deduces the strength of the parameters involved. The analysis compares data extracted from the current video frame against the data obtained from the frame where the eyebrow is in its neutral position.

a) Algorithm procedure: **Binarization**: Eyebrows and skin are separated using an adaptive binarization technique. When studying head anatomy, it can be seen that eyebrows do not present the same aspect in the totality of their arch. The area situated on the inner part of the *Superciliary* arch is generally better defined and easier to differentiate from the skin than the eyebrow arch that goes toward the joint with the *Superior Temporal Line*, because the latter is usually more sparse.

The developed binarization algorithm analyzes the eyebrow in two different zones. Vertical line a-b in Fig. 3 delimits the zones; there, the eyebrow usually changes shape direction and texture. Two different thresholds are applied, one per zone i. Each one is obtained by analyzing the histogram distribution of the corresponding area

$$\text{Th}_i = \min_i + \frac{\max_i - \min_i}{3}. \tag{7}$$

If $\text{Th}_i < pixel\_value$, the pixel is considered part of the eyebrow. The threshold has been chosen to be at a third of the intensity distribution because the analysis area covers three major intensity zones, all of which are well differentiated in most lighting conditions: the eye zone under the eyebrow, the eyebrow itself, and the forehead zone over the eyebrow.

Thinning: We perform a vertical thinning over the binarized image. Under unknown conditions, the eyebrow arch is robustly obtained by detecting the gradient between forehead and eyebrows.

Parameter deduction: The parameters that model eyebrow behavior are deduced by comparing the thinned arch at the current frame against the arch obtained from the analysis of the eyebrow in its neutral position. The process starts by deducting general eyebrow behavior. The median vertical value of the arch (median of the y-component of the points shaping the arch) is compared against the median vertical value of the neutral arch. If the current median is greater than the neutral one, we conclude that we are analyzing upward expressions; otherwise, the downward model representation is used. After selecting the model,

the most significant data from the arch are extracted and used to obtain the model parameters.

III. Applying Temporal and Physical Inter-Feature Constraints to Improve the Analysis

When analyzing facial features from video input recorded in unknown environments, very few assumptions can be made because we have no guarantees regarding quality or specific lighting over the face. To create robust algorithms, the development of our sight expression analysis is based on the following premises: a) the behavior of the features to be analyzed is known and it can be partly modeled; b) the physical structure of eyes and eyebrows is similar in many human beings and image processing algorithms profit from this fact; and c) the features will be assumed to be completely visible, occlusions will be taken as an uncontrolled source of misleading results, as it happens with extreme lighting conditions.

The processing techniques proposed in this section try to globally minimize the influence of unknown sources of error and improve the understanding of sight behavior. As seen in Section II, the motion model of each feature relies on realistic, anatomical and standard motion considerations (intra-feature constraints). The global action understanding of eyes and eyebrows is obtained after applying some inter-feature constraints derived from the natural motion correlation existing between eyes, and among eyes and eyebrows.

A. Applying Inter-Eye Constraints Through a Temporal State Diagram

Synthetic motion generated from the analysis of eye behavior must recreate natural human actions. Unexpected analysis results could lead to unnatural and strange eye motion that would interfere with proper communication. To avoid generating unpleasant eye synthesis, we introduce the following inter-feature constraint: pupil motion is correlated and both eyes behave alike (left [L] & right [R]). This constraint is imposed by filtering each individual state with a Temporal State Diagram, depicted in Fig. 4, whose mission is to assign the most convenient state assuming the same behavior in both eyes and using previous results to compensate for misleading analysis.

B. Eye-Eyebrow Spatial Correlation: Studying Extreme Expressions

Our analysis algorithms perform robustly thanks to their simplicity. This simplicity restricts the understanding of individual motion to natural, coherent eye and eyebrow movements; these constraints are suitable in human-to-human communications but may undesirably filter those details that add strength to expressions.

To partially compensate for this shortcoming, we propose exploiting the existing eye-eyebrow motion correlation to enrich the overall ocular expression understanding from the individual analysis of each feature. When the eyes are closed, the eyelids may behave in two different ways, they may be closed without any tension if the eyebrows are neutral or pulled up; or they may be tensely closed if the eyebrows are pushed down.

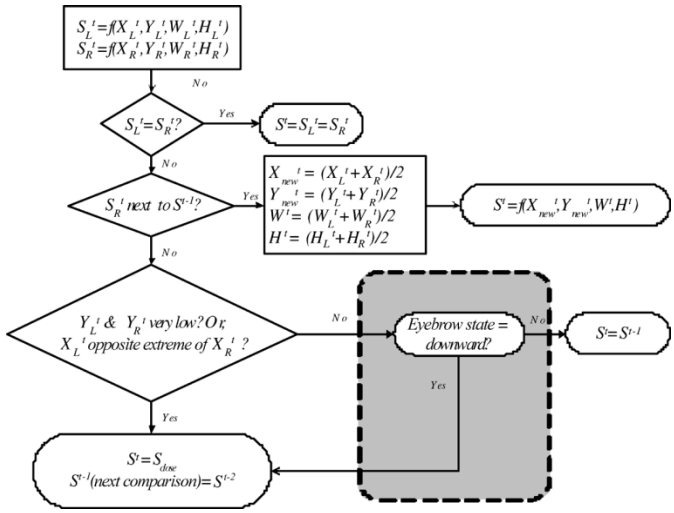

Fig. 4. Analyzed results of each eye feature are filtered through this Temporal Diagram. Current eye states S_L^t and S_R^t are contrasted to obtain a common state for both eyes: S^t. The arrows indicate the procedure flow that leads to eye states. As the state S_{close} does not have any physical information about the pupil location, it is ignored for future analysis in the temporal chain. The starting state is fixed with the X, Y of the eyes in their neutral position.

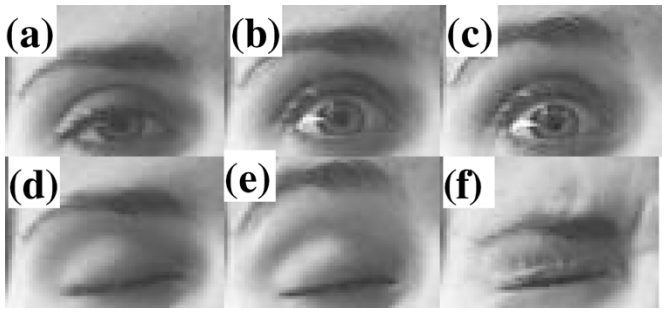

Fig. 5. When the eye is closed (lower row), the eyelid change due to eyebrow action can be used for some specific animation. When the eye is open (upper row) the eyebrow actions must be taken into account to alter the standard vertical motion of the eyelid.

When the eyes are open, the level of the eyebrow height indicates the degree of opening of the eyelid. Fig. 5 illustrates this clear eyelid-eyebrow correlation. Extreme eyebrow actions determine and refine eye motion by:

1) extending the information inside the eye Temporal State Diagram to include the inter-feature constraints derived from eyebrow analysis;
2) deriving the final synthetic eyelid behavior from the position obtained with the pupil location in the eye state analysis plus an extra term accounting for the strength of eyebrow movement, as we present in Section V-C.

IV. Extending Frontal Feature Analysis Algorithms to Any Given Pose

To obtain precise information about a person's pose in space, we have developed an algorithm that utilizes a feedback loop inside a Kalman filter. Kalman filtering has been applied to head tracking giving very positive results [15], [16], and enables the prediction of the translation and rotation parameters of the head

from the 2-D tracking of specific facial features on the image plane.

In [17], we showed how to achieve improvement in the amount of freedom of motion in front of the camera if the speaker's clone is used during the tracking. In our approach, the algorithm operating on the image plane extracts the 2-D features to be tracked on the video sequence from the synthesized image of the model, onto which the predicted pose parameters have already been applied (it is then providing an adjusted view of the user in its future pose).

We have inserted the eye and eyebrow motion analysis algorithms inside this head-tracking framework. For the adaptation process to be possible, the head-pose tracking algorithm and the image processing must share the same observation model. Other performing tracking algorithms work using local 2-D image reference systems on which they can only estimate the user's position on the screen (e.g., Bradski's Cam Shift algorithm [18]). For the feature analysis and the interpretation of the analyzed results, this information is not enough.

In [19], we presented the preliminary results on extending the algorithms to analyze faces in any pose. In this section, we introduce the technique that extends the 'near-to-frontal' analysis algorithms detailed in Section III to any other pose, thus allowing the user freedom of movement when facing the camera.

A. Overview of the Proposed Approach

We first reformulate the motion model and redefine the regions of interest of each feature on an accurate 3-D head representation of the user (simulating the head in front of the camera) in its neutral state (directly facing the camera). Next, we use the pose information to project the ROIs on the video and we apply the image processing required to extract the data for the motion model. Finally, we inverse the projection and the pose transformation of those data to obtain their 3-D equivalent that will be ready to be compared to the motion models already defined in 3-D. Fig. 6 illustrates how this process is applied for the eye state analysis algorithm.

To relate 3-D space, where we relocate the adapted feature motion models, to the 2-D video plane, an observation model is required. Fig. 7 shows the proposed observation model: its reference system and the relationship of the pose transformations with the focal length F of the perspective projection involved. The mathematical transformation associated with this model is as shown in (8), at the bottom of the page, where c_φ stands for $\cos(\varphi)$, s_φ for $\mathrm{in}(\varphi)$, subscript p for projected coordinates and n for neutral 3-D coordinates. This transform reflects the rotations (α, β and γ) around the reference axis and translations (t_X, t_Y and t_Z) along the axis due to the pose. It also contains the perspective projection that needs to be applied to a 3-D synthetic representation of the user's head to recreate the same face image on the video.

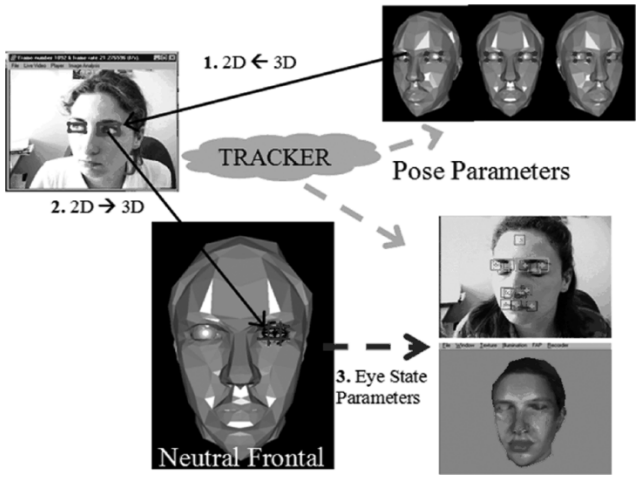

Fig. 6. Diagram illustrating the general adaptation process applied to the eye analysis algorithm; for a better illustration the user's clone has been replaced by a simpler 3-D model. Procedure: 1) prediction of the 2-D ROI; 2) feature image analysis and 3-D inversion; and 3) 3-D interpretation and motion parameter generation. In steps 1 and 3, the pose parameters deduced by the head tracker are used.

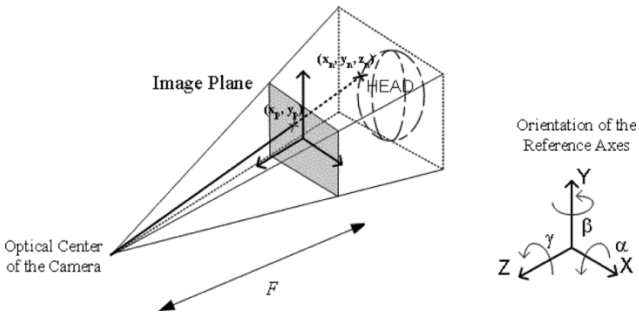

Fig. 7. Reference system and camera model used (of focal length F) for the adaptation process. It establishes the relationship of a point in the Euclidean space $(x_n, y_n, z_n)^{\mathrm{T}}$ and its projected counterpart on the camera image plane $(x_p, y_p)^{\mathrm{T}}$. The axis orientation is such that the camera only sees the negative part of the z-axis.

B. Feature Surface Models for Eye and Eyebrow

System (8) is not a bijective transform; inverting pose and projection generates the straight line that defines the ray of possible solutions in 3-D space for a given projected point. The coordinates of one specific point are the solution to the intersection of this ray with the surface to which the point belongs. We impose the constraint that will enable us to obtain 3-D location of the point by modeling the surface of the feature we are analyzing.

The eye and eyebrow motion analysis algorithms and their parameters have been defined over the image plane. One natural way to extend their use is to reformulate the analysis algorithms on a surface plane whose orientation is parallel to the image plane (perpendicular to the z-axis) and as adjusted as possible to the feature being analyzed.

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \frac{F}{N}\begin{bmatrix} c_\beta c_\gamma x_n - c_\beta s_\gamma y_n + s_\beta z_n + t_X \\ (s_\alpha s_\beta c_\gamma + c_\alpha s_\gamma)x_n - (s_\alpha s_\beta s_\gamma - c_\alpha c_\gamma)y_n - s_\alpha c_\beta z_n + t_Y \end{bmatrix}$$
$$N = (c_\alpha s_\beta c_\gamma - s_\alpha s_\gamma)x_n + (-c_\alpha s_\beta s_\gamma - s_\alpha c_\gamma)y_n - c_\alpha c_\beta z_n - t_z + F \quad (8)$$

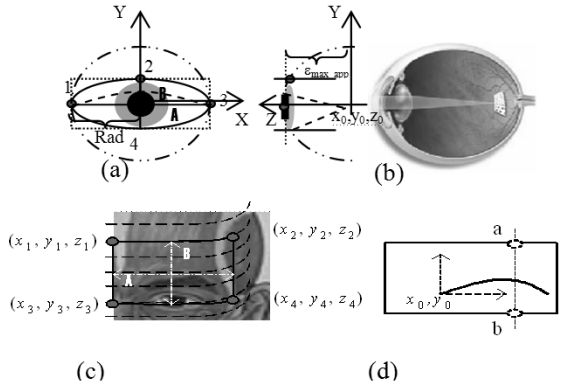

Fig. 8. Eye and eyebrow motion models defined in 3-D space. They are finally located over the 3-D head model used for the analysis, all data needed to define the motion model is obtained from the 3-D head model.

To adjust each feature, we develop the linear approximation of the anatomical surface of the feature on the plane $z_n = M$. This technique permits the straight definition of the 3-D ROI and all the analysis parameters on the surface because the approximation is defined on the neutral model and therefore the plane surface undergoes the same pose transformations as the head. It also ensures a solution by solving the linear system shown in (9), at the bottom of the page.

The development of the mathematical deduction of system (9) from (8) can be found in [20].

We can model the eye as the sphere—$(x_n - x_0)^2 + (y_n - y_0)^2 + (z_n - z_0)^2 = Rad^2$—that best suits the eyeball on the 3-D head representation [Fig. 8(a)]. Then the adjusted plane surface is the linear approximation of the sphere on the point tangent to the pupil in its neutral position: $M = z_0 + Rad$ [Fig. 8(b)]. For the eyebrow, the model is generated from the vertical surface tangent to the forehead at the eyebrow level. The plane approximation is taken from the average value of the depth component of the ROI defined on this surface: $z_n = M = 1/4(z_1 + z_2 + z_3 + z_4)$ [Fig. 8(c)].

C. 3-D Definition of the Region of Interest

One important limitation in performing facial feature analysis is the detection and control of ROI suitable for each algorithm. Usually, this is done manually. Rather inaccurate adaptations of ROIs are done along the time while the speaker is moving. Defining ROIs in 3-D on the speaker's clone, along with the algorithm parameters, helps to deduce automatically suitable 2-D-ROIs on the image plane. The evolution and changes of each feature's ROI on the video sequence are perfectly controlled. Projecting the 3-D regions, frame by frame, on the image automatically reshapes the areas on the frames along with the pose and allows us to foresee if relevant information will be obtained after studying the targeted zone. It is useful to define a threshold, Th, for the computed surface of the projected ROI, under which the algorithm will not perform because it will be considered that there is not enough visible surface. The area expression is

$$\begin{aligned} &Area \\ &\quad = base \cdot height \\ &\quad = \mathbf{A}_{\mathrm{P}} \cdot \mathbf{B}_{\mathrm{P}} \\ &\qquad \cdot \sin\left(\arccos\left(\frac{(x_p^3 - x_p^1)(x_p^2 - x_p^4) + (y_p^3 - y_p^1)(y_p^2 - y_p^4)}{\mathbf{A}_{\mathrm{P}} \cdot \mathbf{B}_{\mathrm{P}}}\right)\right) \end{aligned}$$

where $\mathbf{A}_{\mathrm{P}} = \|3_p - 1_p\|$, $\mathbf{B}_{\mathrm{P}} = \|2_p - 4_p\|$, and $i_p = (x_p^i, y_p^i)$.

D. Theoretical Study of the Extension of the Frontal Analysis to Other Poses

1) Geometrical Interpretation of the Analysis: The circumstances under which the method is unstable are set by determining when the determinant of system (9) is zero

$$F \cdot (F \cdot c_\alpha c_\beta - x_p \cdot s_\beta + y_p \cdot s_\alpha c_\beta) = 0. \tag{10}$$

Analyzing the geometrical nature of system (9), we restrict the study to the case where F is a positive number and vector $\vec{r} = (x_p, y_p, -F)$ is the ray of solutions of the projection inversion (Fig. 7). Then, (10) represents the family of planes that are parallel to this ray, it describes the combination of rigid motion parameters that applied to a plane lead to the nonintersection of vector and plane for a given observed point on the image.

The stability of the system is independent of angle γ, which is the angle that indicates the rotation around z-axis. This is because in our observation model the z-axis and the camera optical axis coincide. The system is unstable for those combinations of angles α and β that transform the plane parallel to the image plane to a plane that contains vector $\vec{r}$.

As the pose of the plane approaches the conditions under which the system becomes unstable, the projected image of the feature being analyzed starts to concentrate toward the same line. The plane is an infinite surface but we are only interested in that part of the plane containing the feature template. Therefore, the instability of the system can also be controlled from the analysis of the ROI projection; as its area decreases, the data on the

$$\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \cdot \begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} a_4 - Ma_3 \\ b_4 - Mb_3 \end{bmatrix} \qquad z_n = M$$

$$\begin{aligned} a_1 &= -x_p \cdot (c_\alpha s_\beta c_\gamma - s_\alpha s_\gamma) + F \cdot c_\beta c_\gamma & b_1 &= -y_p \cdot (c_\alpha s_\beta c_\gamma - s_\alpha s_\gamma) + F \cdot (s_\alpha s_\beta c_\gamma + c_\alpha s_\gamma) \\ a_2 &= x_p \cdot (c_\alpha s_\beta s_\gamma + s_\alpha c_\gamma) - F \cdot c_\beta s_\gamma & b_2 &= y_p \cdot (c_\alpha s_\beta s_\gamma + s_\alpha c_\gamma) + F \cdot (-s_\alpha s_\beta s_\gamma + c_\alpha c_\gamma) \\ a_3 &= x_p \cdot (c_\alpha c_\beta) + F \cdot s_\beta & b_3 &= y_p \cdot (c_\alpha c_\beta) - F \cdot s_\alpha c_\beta \\ a_4 &= -x_p \cdot (t_Z - F) - F \cdot t_X & b_4 &= -y_p \cdot (t_Z - F) - F \cdot t_Y \end{aligned} \tag{9}$$

template plane starts concentrating until it reaches an unstable situation.

2) Precision Error Due to Surface Linearization: This inaccuracy is known and limited

$$|\varepsilon_{\max}| < \arg\max(M - z_i) \quad \text{where } i : \text{point} \in \{\text{feature surface}\}.$$

For instance, for the proposed eye feature model

$$|\varepsilon_{\max\_\text{eye}}| < Rad \ \text{(Fig. (8a))}.$$

Influence of the Pose Determination on the Final Expression Analysis: In [21], the error expressions of the obtained neutral coordinates from the analyzed projected data have been developed. From the analysis of those expressions we conclude that for small inaccuracy errors the overall error behavior shows a clean linear evolution.

Pose parameters related to the x and y-axis (α, β, t_X, and t_Y) behave similarly. Rotations around the x- and y-axis have stronger action over the error results than the rotation related to the z-axis. For given pose conditions γ-rotation, t_X, and t_Y do not change the image appearance of the studied ROI and therefore errors due to inaccuracy in their prediction have less impact on obtaining the neutral coordinates.

The pose-parameter estimation is critical for the template adaptation. In addition to directly influencing the interpretation of the data analyzed on the image, it is also responsible for the success and accuracy of the image processing. The motion template analysis techniques implemented depend on the correct delimitation of the ROI of the feature that is to be analyzed. Since image ROIs are obtained from projections, pose parameters also intervene in determining their final location.

Template adaptation can only be feasible if the rigid motion study of the head on the image generates the required pose parameters with a high degree of accuracy. The level of precision is set so that ROIs are properly tracked and their location does not become the major source of error for the image-processing analysis techniques involved.

V. Experimental Evaluation

The main goals of our experiments are, in one hand, to determine the level of improvement in natural facial-motion data-generation when using intra-feature and inter-feature constraints, and, in the other hand, to evaluate how much freedom of movement the 3-D adaptation allows the speaker.

A. Eye-State Analysis

We used video sequences of 15 Caucasian individuals. The speakers were asked to rigidly stand facing the camera. Each person was asked to move his/her eyes in a natural manner (up, down, left, right, and closing eyes). Three light orientations were used for the recordings. One set of sequences was shot under natural standard conditions in our working space: overhead fluorescent lighting. Another set of sequences was recorded under extreme lighting conditions: direct light coming either from the front, from the right or the left side. The average length of each sequence was of 500 frames at 20 frames/s, and the average eye ROI was 32 × 24 pixels.

TABLE I
Test Results on Eye-State Analysis

SEQUENCES:			O-H FL. LIGHT	FRONTAL	LEFT SIDE	RIGHT SIDE
PERCENTAGE	WITHOUT	O - 3%	16.84	17.10	27.77	9.53
		O - 5%	34.95	43.33	38.15	23.26
		O - 10%	73.47	68.15	90.29	86.28
	WITH STATE DIAGRAM	C - 3%	88.78	72.60	79.91	85.35
		C - 5%	95.15	84.78	88.94	78.60
		C - 10%	93.88	86.65	90.97	86.29
		C_{CE} - 3%	4.68	0.32	0.00	0.00
		C_{CE} - 5%	9.26	13.54	0.00	0.00
		C_{CE} - 10%	3.59	17.30	5.65	0.00
		E - 3%	27.27	28.21	62.92	73.02
		E - 5%	0.00	18.46	61.22	33.70
		E - 10%	0.00	21.05	2.50	37.63

To consider that the Pupil Search Algorithm was successful, we used the following criteria:

— quantitatively, the X and the Y components of both eyes were the same, computed with the following intervals of accuracy:

$$|X_L - X_R| \leq 3\%, 5\% \ or \ 10\% \ of \ W_{\text{ROI}}$$
$$|Y_L - Y_R| \leq 3\%, 5\% \ or \ 10\% \ of \ H_{\text{ROI}}$$

— qualitatively, the result obtained defined the expected eye action. This was visually inspected.

To understand the level of improvement achieved when adding the Temporal State Diagram, we compared the results obtained not using the Temporal Diagram (Table I-O) against those obtained after utilizing the diagram (Table I-C). The improvement is significant; above all, the state diagram could determine the S_{closed} state that the algorithms utilized independently cannot. Lighting conditions influence the image analysis results but the state diagram ensures a success rate of 70%–90% by providing animation data that is as smooth as possible and that generates natural eye motion. The most disturbing effects come from the introduction of S_{closed} where there has not been any such action; Table I plots the percentage of errors coming from detecting a false S_{closed} state (Table I-E). In (Table I-C_{CE}) we count the percentage of correct analyzed closed-eye actions (that will be rendered to a closed eye) without the need of being detected as S_{closed} in the diagram. We conclude from the analysis that a tradeoff between the accuracy of motion and the robustness of the Temporal State Diagram is needed. The larger the interval of accuracy permitted, the smaller the margin of action that the Temporal State Diagram will have for correcting errors and consequently, the less smooth the eye motion will be. Fig. 9 illustrates the adjustment in gaze motion after filtering the analysis results with the Temporal State Diagram.

None of the speakers wore glasses for these tests. Side experiments were made to evaluate what was the behavior of the system to wearing glasses. The empirical results showed that

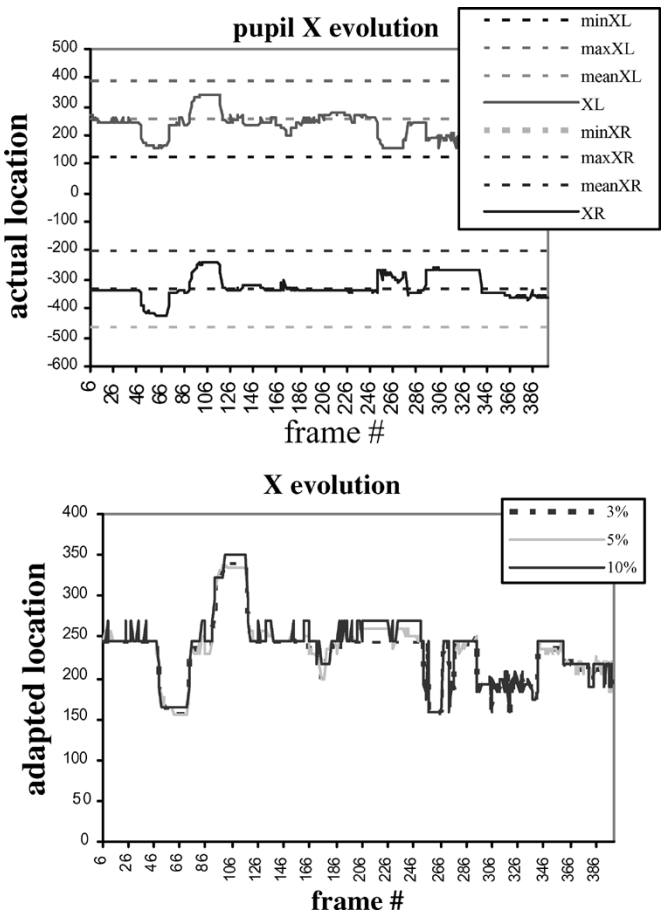

Fig. 9. Upper graph: evolution of the extracted data regarding the pupil X location for both eyes. Lower graph: resulting X location after applying the Temporal State Diagram. It shows the results for $f(X, Y, W, H)$ that quantizes with an accuracy of 3%, 5%, and 10% of W_{ROI} (example sequence: "O-H. FL. LIGHT").

glasses were not an obstacle as long as they did not reflect light. In general, occlusions (even if they come from glasses) are treated as errors. The best-effort nature of the system ensures that natural motion will be generated if occlusions occur.

B. Eyebrow Analysis

To check the correct behavior of the model and its application for eyebrow motion analysis, we have applied the binarization-thinning technique over the same eye on the frames of several video sequences of individuals of different eyebrow characteristics recorded under uncontrolled lighting conditions. To verify that the parameters obtained actually correspond to the eyebrow behavior, we have compared the thinning results of each frame against the obtained arch by applying the model over the neutral position arch (this result is called modeled arch).

The best way to evaluate the performance of our techniques is to visually compare arch results; unfortunately, this process is not suitable to be applied to large amount of data.

To interpret the performance correctness of our approach, we have defined two different measurements.

— A *pseudoarea*: $\widetilde{a} = \sum_i |y_k[i] - y_m[i]|$, which can be understood as the area contained in-between arch k and m, and it denotes the shape similarity between them; the closer $\widetilde{a}$ is to 0, the more alike they are. We compare the area difference between the neutral arch and the current frame arch against the shape difference between the modeled arch and the current frame arch. This measurement checks if the eye shape modeled by the extracted motion parameters follows the expected eyebrow obtained from the action.

— A *mean difference comparison*: where we compare the mean (average vertical value of the arch) difference between the current frame arch and the neutral position arch against the mean difference between the current frame arch and the modeled arch. This information helps us to evaluate why the analysis procedure was not able to detect the right eyebrow's general action.

We consider that the algorithm has worked correctly if the pseudoarea comparison shows that the modeled arch is closer to the current frame arch than the neutral position arch.

Visually inspecting the results of our experiments, we conclude that accuracy errors mostly come from the image processing performance of the analysis rather than from the motion model used. The percentage of estimation success (better measurements over the modeled arch) is around 85% for those sequences where image quality and environment lighting conditions are standard. For video input recorded under extreme lighting conditions, performance drops to around 70%. We must point out that the worst estimation usually happens for low expression movements, where the inaccuracy of the location of the analysis area (the speaker may slightly move) is large enough to distort the average results. In this case, as the average difference measurement shows (Fig. 10), we may interpret an up movement as being down or vice versa.

C. Eye-Eyebrow Cooperation

During the eye-eyebrow cooperation tests, we have compared the evolution of the vertical eyelid motion derived from the eye-state analysis against the results obtained after adding the influence of the eyebrow

$$y_{\mathrm{eyelid}}^{\mathrm{final}} = y_{\mathrm{eyelid}} + \mu' \cdot \mathrm{FAP} \text{ with } \mu' = \frac{\left(y_{\mathrm{eyelid}_{\mathrm{MAX}}} - y_{\mathrm{eyelid}}\right)}{\mathrm{FAP}_{\mathrm{MAX}}}.$$

The new eyelid up-vertical motion, $y_{\mathrm{eyelid}}^{\mathrm{final}}$, is the sum of its original obtained value, y_{eyelid}, plus a term proportional to the eyebrow FAP magnitude (ranged between 0 and $\mathrm{FAP}_{\mathrm{MAX}}$). μ', the proportion coefficient, is dependent on the maximum of that range, $\mathrm{FAP}_{\mathrm{MAX}}$, the maximum value for the eyelid motion, $y_{\mathrm{eyelid}_{\mathrm{MAX}}}$, and y_{eyelid}. FAP stands for Face Animation Parameter [13] and, in this context, refers to the magnitude of the parameter that controls the synthetic vertical movement of the eyebrow.

Looking at Fig. 11, we can clearly see the existing correlation between eyelid and eyebrow motion. The shadowed parts highlight the frames where the person was closing his eyes. The FAP evolution, which depicts the action of the eyebrows, shows that with eyebrows moving up the eyes were open (Fig. 5(c) is taken from frame 221 of sequence showed in Fig. 11 and Fig. 12), and when eyebrows where moving down, closed eyes were detected (Fig. 5(f) is taken from frame 301 of the same sequence). In

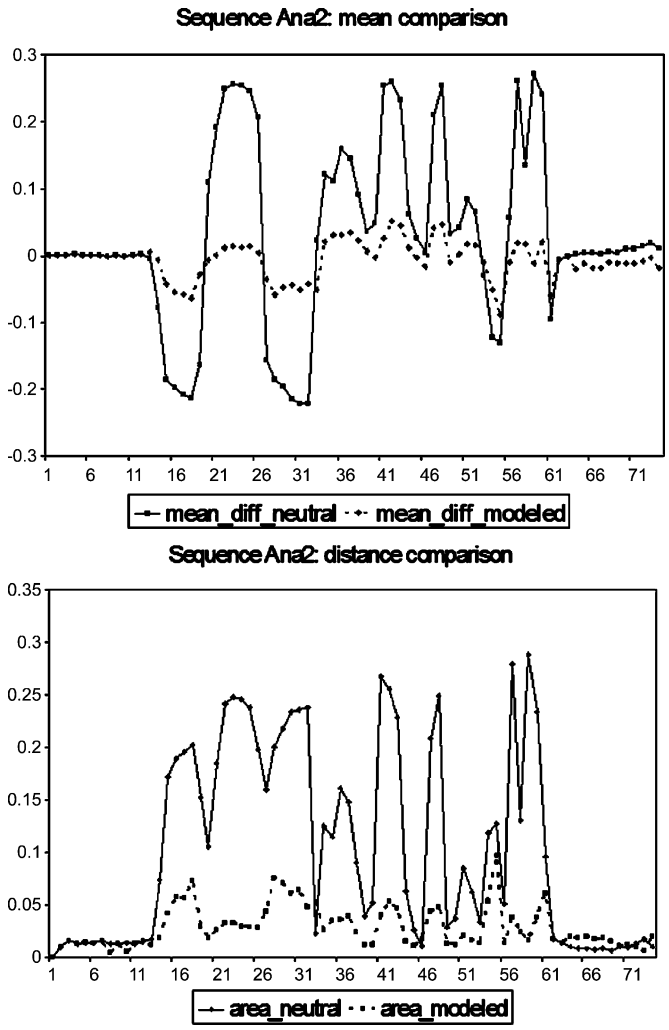

Fig. 10. The algorithm detects the right movement (the mean difference decreases) and estimates the motion parameters correctly (the area decreases). We observe the best behavior for extreme eyebrow expressions. Detailed results of our experiments are presented in [21].

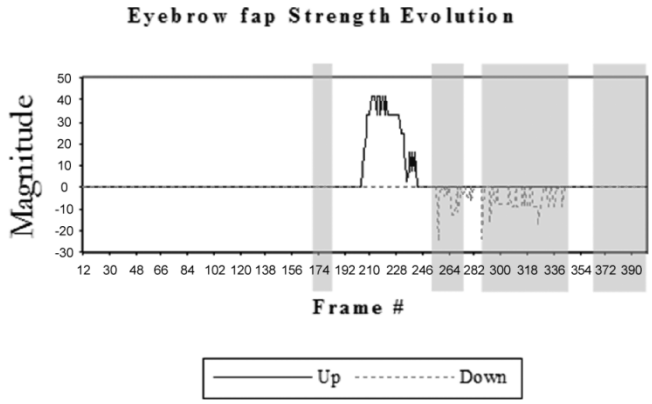

Fig. 11. Eyebrow magnitude evolution (0-$\mathrm{FAP}_{\mathrm{MAX}}$) taken from sequence "O-H. FL. LIGHT". The shadowed areas indicate when the eyes are closed.

Fig. 11 and Fig. 12, we prove that the added term to eyelid motion correctly improves the motion strength without interfering with the eye analysis when no eyebrow motion is detected.

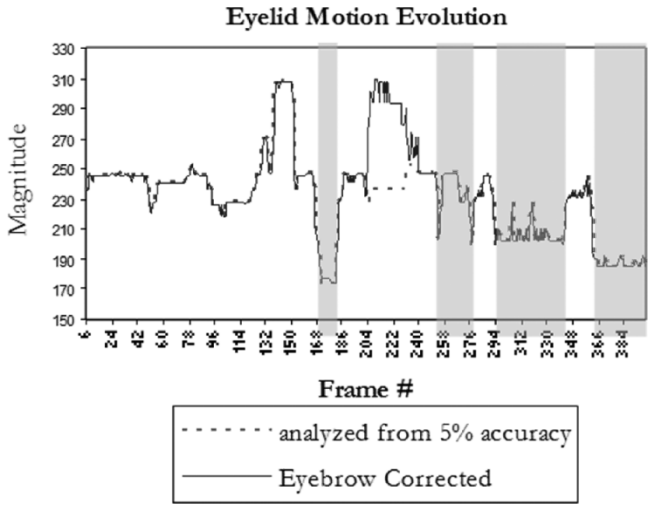

Fig. 12. Eyelid standard analysis results compared to the corrected results after weighing the analysis with eyebrow data. Analysis made on sequence "O-H. FL. LIGHT". The shadowed areas indicate when the eyes are closed.

Fig. 13. Synthetic images and video input are blended to be able to initialize the analysis system.

D. Testing the Adaptation

To evaluate the expression/pose-tracking coupling technique we used a highly realistic 3-D model (clone) obtained from scanning the speaker's head with a Cyberware scanner. The ROIs and other parameters of the algorithms were defined manually beforehand on the planes that approximate the 3-D surfaces of the features. These tests do not evaluate the accuracy of the head model; 3-D data is considered ground truth. Tests were performed using sequences of one individual and his/her clone. To start every test, speaker and 3-D model were visually aligned in a frontal position (see Fig. 13). At that point, the neutral state for all features was determined. This was done once per sequence or live analysis.

1) Interference in the Image-Processing: Deformation of the ROI and Introduction of Artifacts From Other Features: The first interference in the image-processing analysis comes from

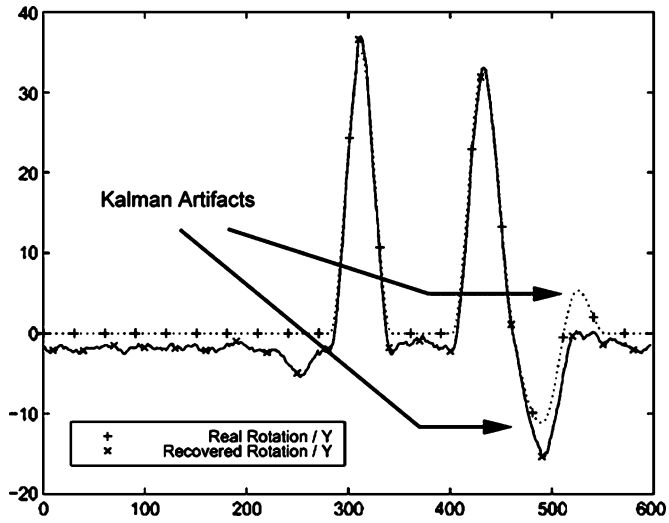

Fig. 14. Real and recovered Y position of a sample sequence.

the adaptation of the theoretical ROIs to the physical square nature of frames in video memory. For practical purposes, to make the deformed areas more suitable for image analysis, we enclose them in video analysis rectangles.

This ensures that the ROI and its feature are completely enclosed inside the analyzed area. Unfortunately, it also implies the inclusion of some artifacts coming from other facial features next to the one being analyzed. In some cases, when analyzing eyebrows (with hair and eyes), this can be taken into account during image processing; otherwise, they will be possible sources of error that the system will have to control.

2) Evaluating the Influence of the Dynamics of the Kalman Filter-Based Pose Tracking: In [17], a complete and exhaustive analysis of the behavior of the extended Kalman Filter of the utilized head pose tracker has been developed. From this analysis, we would like to point out than even when the tracker works fine, the obtained results are slightly noisy and the strongest artifacts appear in the presence of rapid transitions, thus indicating that the prediction model of the Kalman filter, assuming constant acceleration, is not always appropriate (see Fig. 14).

The interference of the noise of the predicted pose parameters in the feature motion interpretation is comparable to the image-processing accuracy. Strong tracking artifacts, which are not common, mask any other effect and might mislead the results.

3) Evaluation of the Robustness of the Eye and Eyebrow Motion Models Adapted to Other Poses: To study the performance of the coupling in live input, we have developed some applications that integrate the eye and eyebrow analysis inside the pose tracking system. These applications analyze recorded sequences and direct camera input. Analysis rate was around 1.5 frames/s for recorded video input whose size was 384 $\times$ 288 and whose acquisition rate had been 30 frames/s. If direct live input was used, frame rates increased to 5 to 10 frame/s because no decoding of video sequences was involved. The developed code has not been optimized and the highest computational restriction comes from the block matching required for the head tracking, which we are aware can be improved. Tests have been performed in uncontrolled lab conditions as well as in different unknown environments [22].

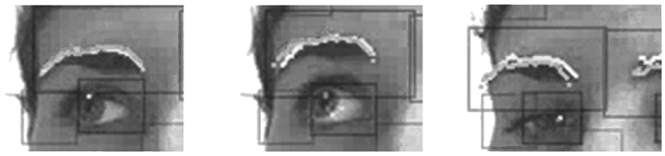

Fig. 15. The right eyebrow has been extracted from one sequence. We observe the evolution of the head rotation and at the same time as the eyebrow moves upwards. The white line represents the arch extracted from the image processing analysis; the grey line is the result from the projection of the neutral motion model after having applied on it the motion parameters. Ideally, both arches should have the same shape and location. The pupil tracking from the right eye is also plotted. The rectangles are the eye and eyebrow ROIs and the squares are the blocks utilized during the pose tracking.

The success of the analysis was visually verified. We used an animated 3-D head model to instantly replicate the eye motion of the speaker. For the eyebrows, the feedback evolution was obtained comparing the analyzed arch with the modeled arch over the input video images for different head movements. (Fig. 15). The tests we recorded in some video sequences [23]. They show that, as foreseen in the theoretical study, even when the tracking is steady and accurate, the surface modeling starts distorting the results for angles α and β greater than $\pm\pi/4$. This is basically because surface modeling with plane surfaces cannot take into account the partial hiding of the modeled features.

VI. Conclusion

Ocular expressions are crucial for communications and little taken into account during the development of this kind of applications. Although there are many techniques for eye analysis [24], very few have been identified that study eyebrow and we have not found any that analyzes both together, exploiting their natural correlation. The experimental results obtained from our framework for eye-eyebrow motion understanding are very positive and encourage the authors to continue developing the same strategy to analyze other facial features: mouth, wrinkles... whose analysis and modeling seem a priori more complex. In our future research line, we intend to introduce image analysis techniques with customized parameters that will adapt to the user characteristics (race, gender, skin tone, etc.). This will be achieved through the use of large training sets that will help to adjust the system parameters to any circumstances. Taking advantage of these training sets the current inter- intra-feature constraints will be also customized based upon the user racial characteristics.

A future challenge will appear to couple all facial feature analysis with the proposed Kalman-based pose tracking system. As more features may move, fewer facial tracking points will be available for the head tracking algorithm.

At that point, studies on the robustness of pose tracking versus the number of features and the freedom of movement will have to be done and even the search for a possible head tracker substitute could be envisaged, for instance, a speaker-independent head tracker. Needless to say, this new tracker would need to maintain the accuracy performance requirements of the complete pose-expression system.

With the use of synthetic facial reproduction, it is possible to interfere with reality by improving the final visual representation during communications, e.g., by adapting the user's point

of view of the speaker. Although the technique described in this paper has been addressed as a solution to analyze ocular expressions to obtain animation parameters for synthetic facial motion, it can also be extended to other scientific fields where the knowledge of instant actions from people in front of cameras is desired [25], for instance, in Human Computer Interaction analysis [26]. Unlike other techniques applied for source coding, we provide the means for understanding of emotion.

REFERENCES

[1] P. Eisert, T. Wiegand, and B. Girod, "Model-aided coding: a new approach to incorporate facial animation into motion-compensated video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 3, pp. 344–358, Apr. 2000.

[2] L. Yin and S. Yu, "Low bit-rate video-conferencing scheme compatible with CCITT Rec.H.261," *J. China Inst. Commun.*, vol. 14, no. 2, pp. 3–10, 1993.

[3] J.-L. Dugelay, K. Fintzel, and S. Valente, "Synthetic natural hybrid video processings for virtual teleconferencing systems," in *Proc. IEEE Picture Coding Symp.*, Portland, OR, Apr. 1999.

[4] S. Redfern and N. Naughton, "Collaborative virtual environments to support communication and community in internet-based distance education," *J. Inform. Technol. Educ.*, vol. 1, no. 3, pp. 201–211, 2002.

[5] M. Garau, M. Slater, V. Vinayagamoorhty, A. Brogni, A. Steed, and M. A. Sasse, "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment," in *Proc. SIG-CHI Conf. Human Factors in Computing Systems*, vol. 1, 2003, pp. 525–536.

[6] F. Pighin, R. Szeliski, and D. H. Salesin, "Modeling and animating realistic faces from images," *Int. J. Comput. Vis.*, vol. 50, no. 2, pp. 143–169, 2002.

[7] T. Goto, S. Kshirsagar, and N. Magnenat-Thalmann, "Automatic face cloning and animation," *IEEE Signal Process. Mag.*, vol. 18, no. 3, pp. 17–25, May 2001.

[8] M. Kampmann, "Automatic 3-D face model adaptation for model-based coding of videophone sequences," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 3, pp. 172–182, Mar. 2002.

[9] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 9–115, Feb. 2001.

[10] Y.-J. Chang, C.-C. Chen, J.-C. Chou, and Y.-C. Chen, "Virtual talk: a model-based virtual phone using layered audio-visual integration," in *Proc. IEEE Int. Conf. Multimedia and Expo*, New York, NY, 2000.

[11] A. C. Andrés del Valle and J.-L. Dugelay, "Eye state tracking for face cloning," in *Proc. IEEE Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001.

[12] (2001) Anatomy of Eyebrows. [Online] Available: http://acuplastic.co.uk/eyebrows/anatomy.html

[13] A. C. Andrés del Valle and J.-L. Dugelay, "Eyebrow movement analysis over real-time video sequences for synthetic representation," in *Lecture Notes in Computer Science*: Springer, 2002, vol. 2492, pp. 213–225.

[14] S. Valente and J.-L. Dugelay, "Face tracking and realistic animations for telecommunicant clones," *IEEE Multimedia Mag.*, vol. 7, no. 1, pp. 34–43, Jan.–Mar. 2000.

[15] M. D. Cordea, E. M. Petriu, N. D. Georganas, D. C. Petriu, and T. E. Whalen, "3D head pose recovery for interactive virtual reality avatars," in *Proc. IEEE Instrumentation and Measurement Technology Conf.*, Brunswick, ME, Jul.-Aug. 2001.

[16] J. Ström, "Model-based real-time head tracking," *Eurasip J. Appl. Signal Process.*, no. 10, pp. 1039–1052, Oct. 2002.

[17] S. Valente and J.-L. Dugelay, "A visual analysis/synthesis feedback loop for accurate face tracking," *Signal Process.: Image Commun.*, vol. 16, no. 6, pp. 585–608, 2001.

[18] G. R. Bradski, "Computer vision face tracking as a component of a perceptual user interface," in *Proc. IEEE Workshop on Applications of Computer Vision*, 1998, pp. 214–219.

[19] A. C. Andrés del Valle and J.-L. Dugelay, "Facial expression analysis robust to 3D head pose motion," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Lausanne, Switzerland, Aug. 2002.

[20] A. C. Andrés del Valle, "Facial Motion Analysis on Monocular Images for Telecom Applications: Coupling Expression and Pose Understanding," Ph.D. dissertation, Institut Eurécom, ENST Paris, Paris, France, 2003.

[21] A. C. Andrés del Valle and J.-L. Dugelay, "Pose Coupling With Eye Movement Tracking,", Eurecom Res. Rep. RR-01-058, Dec. 2001.

[22] ——, "Online face expression analysis: coupling head-pose tracking with face expression analysis," in *Technical Demo at ACM Multimedia*, Dec. 2002.

[23] Video Recordings From the Tests Performed in [20]. [Online] Available: http:\\MM549.eurecom.fr

[24] A. Haro, M. Flickner, and I. Essa, "Detecting and tracking eyes by using their physiological properties, dynamic, and appearance," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Hilton Head, SC, Jun. 2000.

[25] Eagleeyes Project. (2004) Camera Mouse. Boston College, Boston, MA. [Online] Available: http://www.bc.edu/schools/csom/eagleeyes/cameramouse/

[26] A. C. Andrés del Valle and J.-L. Dugelay, "Making machines understand facial motion and expressions like humans do," in *Proc. Int. Conf. Human—Computer Interaction*, 2003.

Ana C. Andrés del Valle (M'04) was born in Barcelona, Spain, in 1976. She received the M.Sc. degree in telecommunications from the Technical University of Catalonia, Barcelona, in 1999 and the Ph.D. degree in image and signal processing from Télécom Paris, Paris, France, in 2003.

She is currently a Research Specialist at the Accenture Technology Labs, Sophia Antipolis, France, involved in the Home Intelligent Services initiative. Prior to this position, she was a Project Leader at the VICOMTech Research Center, Spain. During her Ph.D. studies, she was a Research Engineer at the Eurecom Institut, France, and held an Invited Professor Fellowship from the University of the Balearic Islands, Spain. She began her work in research after being an Intern for AT&T Labs-Research, Red Bank. She publishes regularly and has contributed to several books, such as the Encyclopedia of Information Science and Technology (Hershey, PA: Idea Group, 2005).

Dr. Andrés del Valle was awarded the "2ème Prix de la Meilleure Thèse Telecom Valley" for the outstanding contributions of her Ph.D. work.

Jean-Luc Dugelay (M'94–SM'02) received the Ph.D. degree from the Department of Advanced Image Coding and Processing, France Telecom Research, Rennes, France, in 1992, where he worked on stereoscopic television and three-dimensional motion estimation.

He joined the Eurecom Institute, Sophia Antipolis, France, in 1992, where he is currently a Professor in charge of image and video research and teaching activities inside the Multimedia Communications Department. His group is currently involved in several national and European projects related to image processing. He also serves as a Consultant for several major companies, including France Telecom R&D. His research interests are in the area of multimedia signal processing and communications, including security imaging (i.e., watermarking and biometrics), image/video coding, facial image analysis, virtual imaging, face cloning, and talking heads.

Dr. Dugelay is currently an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA. He is a member of the IEEE Signal Processing Society and the Image and Multidimensional Signal Processing Technical Committee (IEEE IMDSP TC).