Eyebrow Movement Analysis over Real-time Video Sequences for Synthetic Representation

Ana C. Andrés del Valle and Jean-Luc Dugelay

Institut Eurécom. Multimedia Communications Department. 2229, route des Crêtes
B.P. 193 – 06904 Sophia Antipolis cedex – France
Tel: +33 (0)4.93.00.{29.21}{26.41}; Fax: +33 (0)4.93.00.26.27
{ana.andres, jean-luc.dugelay}@eurecom.fr
http://www.eurecom.fr/~image

Abstract. To study eyebrow behavior from video sequences we utilize a new image analysis technique based on an anatomical-mathematical motion model. This technique 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. The present technique is meant to suit real-time requirements and to work under uncontrolled environments: without specific known lighting or special markers on the face, by using only computer vision and image processing methods.

1 Introduction

Face animation has become a need for those multimedia applications where human interaction with virtual and augmented environments enhances the interface. It is also a solution for face image transmission in low bit rate communications, video-telephony, virtual teleconferencing, etc.

To synthesize the speaker's face expression in real-time communications, we must develop video image analysis techniques robust to any environment, lighting conditions by not imposing any physical constraints. Most of the existing face expression analysis techniques are not really suitable for practical purposes. The most performing analysis systems utilize the synthesis of head models. DeCarlo and Metaxas [1] have developed a technique based on optical flow constraints to analyze and interpret face motion. Eisert and Girod [2] use a similar approach in their teleconferencing system. Their research results are very encouraging but their analysis algorithms work under many environmental restrictions.

In our previous work about virtual teleconferencing environments [3, 4], we have developed a video analysis framework where we artificially separate pose tracking from face expression analysis. To understand face expression motion, we develop some dedicated analysis techniques to apply on eyes, eyebrows and mouth. Our analysis algorithms generate face animation parameters used to synthesize feature motion in 3D. Other approaches, e. g. Y. Tian et al. [5], study more face features; the given solutions are usually too complex to be utilized in real-time. T. Goto at al.

describe in [6] a complete real-time system also based on different feature image analysis techniques. They give few details about the image processing algorithms involved in the analysis of the eyebrows. In general, we have not found in the literature a formal description of analysis algorithms to study eyebrow behavior, even though eyebrows play an active role in emotional face expression.

We utilize a two step process to develop our image analysis algorithms. First, we design image-processing techniques to study the features extracted from a frontal view of the face. Faces show most of their expression information under this pose and this allows us to verify the correct performance of the image processing involved. Second, we adapt our algorithms to analyze features taken at any given pose. To do so, we use the pose parameters predicted during face tracking and the 3D model synthesis of the speaker's clone. This approach has already been successfully tested to analyze the eye feature [7].

To study the eyebrow behavior we have developed an anatomical-mathematical motion model that can easily be used to derive the parameters describing 2D (vertical and horizontal) eyebrow movements. Our technique is here tested for a frontal view and works under no environmental constraint, besides fixing the head pose. As briefly explained in Section 4, in our future perspective, we will adapt the algorithm to couple pose and expression information.

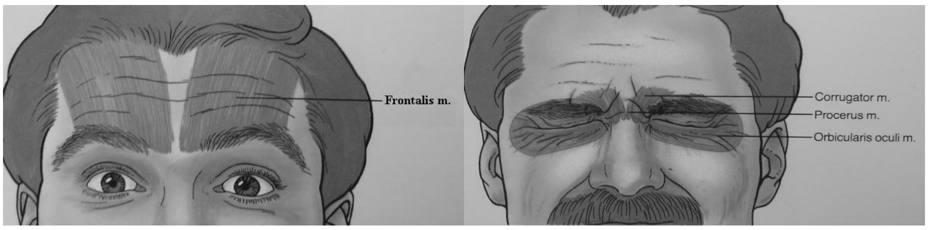

Fig. 1. Several muscles generate the eyebrow movements. Upward motion is mainly due to the Frontalis muscle and downward motion is due to the Corrugator, the Procerus and the Orbicularis Oculi muscles[1]

Table 1. Notation conventions used in formulae

x_n, y_n : real coordinate values of the eyebrow in their neutral position.
$x_n[\,i\,]$, $y_n[\,i\,]$: coordinate value of the pixel obtained from the video image analysis of the eyebrow in its neutral position at position i.
x, y : real coordinate values of the eyebrow in their current (analyzed frame) position.
$x[\,i\,]$, $y[\,i\,]$: coordinate value of the pixel obtained from the video image analysis of the current frame eyebrow at position i.
Δx, Δy : real coordinate difference between the current eyebrow arch and the neutral arch $x_{frame}-x_{neutral}$, $y_{frame}-y_{neutral}$, respectively.
$\Delta x[\,i\,]$, $\Delta y[\,i\,]$: real coordinate difference between the pixels from the current eyebrow arch being analyzed and those from the neutral arch at position i.
[0], [N] and [max] indicate the computed values for the first point ($x = 0$), the last point (x = last) and the point with the maximum vertical value ($x : y = \max$) on the arch.

[1] Images and information based on data from www.oculoplastic.co.uk/eyebrows/anatomy.html

2 Anatomical-Mathematical Eyebrow Movement Modeling

To model the eyebrow movement, we define some mathematical expressions that superficially follow the muscular behavior and interaction when eyebrow motion exists.

Basically, four muscles control the eyebrow movement:
(i) *Frontalis* (F): that elevates them.
(ii) *Corrugatori* (CS): that pulls them downwardly, produces vertical glabellar wrinkles.
(iii) *Procerus*: that lowers the eyebrows downwardly.
(iv) *Orbicucularis oculi*(OO): that closes eyes and lowers eyebrows.

Although the shape is slightly dependent on the anatomy of the person, eyebrow motion is more general. This enables us to represent eyebrows as arches, whose shape is specific to the person but whose motion can be mathematically modeled. We parameterize the arch movement as the displacement in the x, y and z-axis of each of its points compared to the initial neutral position (when no force acts). ($\Delta x = x_{frame} - x_{neutral}$, $\Delta y = y_{frame} - y_{neutral}$ and $\Delta z = z_{frame} - z_{neutral}$). All notation conventions can be found in Table 1.

Two different behaviors exist in eyebrow motion, one when the expression goes upwards and another when the expression goes downwards. Different muscular action is involved for each of them and therefore different models control them. These expressions have been derived from the observation of the muscular motion of the eyebrows and the empirical study of the optical flow behavior of the eyebrow area observed on real video sequences and adapting the parameters involved to the anatomical shape of the eyebrow.

Eyebrow Motion Expressions:

– UPWARDS:

$$\Delta x = Ff_x \cdot e^{(-x_n/\alpha)} \tag{1}$$

$$\Delta y = Ff_y + Ff_y' \cdot e^{(-x_n/\alpha)} \tag{2}$$

– DOWNWARDS:

$$\Delta y = -Fcs_y - Foo_y \cdot (|x_n - \beta| - \beta)^2 \tag{3}$$

If $x_n < \beta$ $\qquad \Delta x = -Fcs_x$ **(4a)**

If $x_n > \beta$ $\qquad \Delta x = Foo_x \cdot (\beta - x_n) - Fcs_x$ **(4b)**

Ff, *Fcs* and *Foo* are the magnitudes associated to the force of the Frontalis muscle, Corrugator muscle and the Orbicucularis Oculi respectively. The action of the Procerus muscle, being close and highly correlated to the one from the Corrugator, is included in the *Fcs* term. $_x$ and $_y$ indicate the different components,

$\alpha = 2 \cdot \frac{w}{3}$ and $\beta = \frac{w}{2}$. All coordinates relate to the eyebrow local coordinate system. Figure 2 depicts the coordinate axis for the left eyebrow; the right eyebrow is symmetrical over an imaginary vertical line between the eyebrows.

$\Delta z = f(\Delta x, \Delta y)$, quite more difficult to model, does not provide critical information regarding the eyebrow expression. Δz is never well estimated from a frontal point of view. If we want to realistically synthesize 3D eyebrow motion with information obtained from image analysis under these conditions, we may estimate the Δz movement by assuming that the eyebrow motion follows the front surface, thus, simulating its natural behavior. The *"displacement of texture coordinate"* synthetic animation technique described in [3] illustrates this concept. This procedure simulates the eyebrow skin sliding motion on the skull. Changing the texture coordinates vertically and horizontally generates the animation; the model 3D coordinates remain unchanged, thus leaving the skull shape untouched.

Applying the formulae over two rounded arches with different parameter values, we obtain deformations that correspond to the expected eyebrow deformation due to those forces. Figure 3 shows the simulation of extreme modeled movement (*downwards* and *upwards*) on both eyebrows.

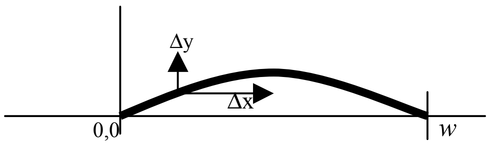

Fig. 2. Eyebrow model arch for the left eye and its coordinate reference. The origin for the analysis algorithm it is always situated at the inner extreme of the eyebrow (close to the nose) and defined for the eyebrow in its neutral state

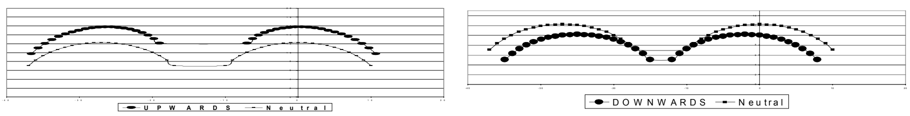

Fig. 3. The action of the eyebrow behavior model applied over the neutral arch results on a smooth deformation. The graph on the left depicts eyebrow rising motion (*upwards*) for positive values of Ff_x, Ff_y and Ff_y'. The graph on the right represents eyebrow frowning (*downwards*) for positive values of Fcs_x, Foo_x, Fcs_y and Foo_y

3 Image Analysis Algorithm: Deducing Model Parameters in a Frontal Position

We achieve two goals by modeling the eyebrow movement. On the one hand, we simplify the eyebrow motion understanding to a point where we can derive its movements on images by comparing image data with the model parameters. On the other hand, this model is complete enough to generate the required information to create synthetic eyebrow expressions.

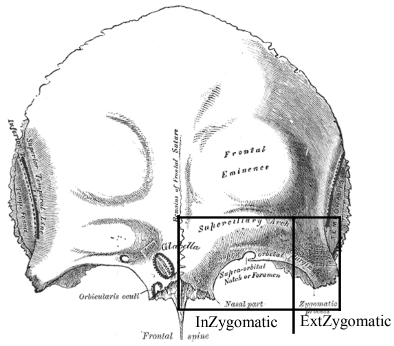

Fig. 4. The eyebrow changes its hair density as it goes away from the inner extreme. The bone structure of the skull determines the shading difference along the eyebrow. We set two different binarization thresholds: Th1 for the *InZygomatic* zone and Th2 for the *ExtZygomatic*

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 a neutral position or *neutral frame*.

Algorithm procedure:

- **Binarization:** Eyebrows and skin are normally easy to separate in terms of hue, and with less accuracy, intensity. Under ‘regular’ although 'not controlled' lighting conditions we can differentiate eyebrows from skin and therefore binarize the feature image. We consider 'normal, not controlled' lighting any illumination over the face that permits the eyebrow visual differentiation on the video frame.

Due to the anatomical nature of the head, eyebrows do not present the same aspect all over 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 towards the joint with the *Superior Temporal Line*, because this last one is usually more sparse. Our analysis algorithm needs to detect the complete eyebrow because we are interested in studying the overall shape behavior. We have developed a binarization algorithm that analyzes the eyebrow in two different zones. One zone includes from the center of the face up to the point which is half way between the *Foramen* and the *Zygomatic* process (point where the eyebrow usually changes shape direction and texture) and the other zone goes from there to the external end of the eyebrow. We refer to Figure 4 to locate the different parts.

To perform the binarization we apply two different thresholds, one per zone. Each threshold is obtained by analyzing the histogram distribution of the corresponding area. The eyebrow is taken as the darkest part on the video image being analyzed.

$$Th_i = \min_i + \frac{\max_i - \min_i}{3} \tag{5}$$

If pixel_value<Th_i, the pixel is considered as 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, which are well differentiated in most lighting conditions: the eye zone under the eyebrow, the eyebrow itself and the front zone over the eyebrow. Usually the darkest one belongs to the eyebrow. Figure 5-a shows the histogram of one of the zones, on which we have marked the three different zones.

– Results of the Binarization Process:

This binarization algorithm correctly detects the eyebrow even if in some cases it also introduces some artifacts. Eyes and hair are often labeled as being part of the eyebrow (see Figure 5-b, where eye is marked as eyebrow). Both, eye and hair must not be taken into account when analyzing the Region of Interest (ROI) to extract the eyebrow arch. Due to their predefined and fixed situation on the ROI, we have easily adapted the thinning algorithm to avoid the possible expected artifacts.

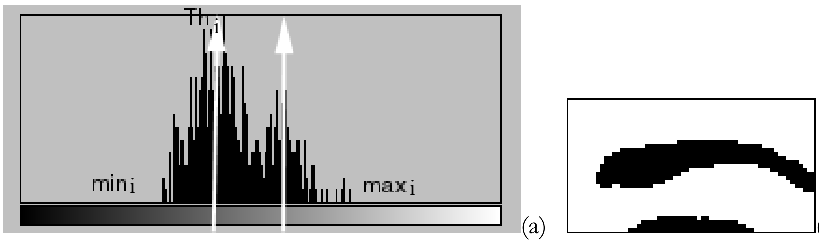

Fig. 5. The eyebrow 'two part' binarization leads to good determination of the eyebrow area but it also may introduce artifacts by labeling eyes or hair as part of the eyebrow. In the current eyebrow binary image we see how the eye has also been detected

- **Thinning:** We perform a vertical thinning over the binarized image to obtain the rounded arch that will define the eyebrow. This vertical thinning delineates the eyebrow arch by choosing the pixels located in the vertical middle of the binarized eyebrow shape. Figures 7 and 8 show some thinning results.

- **Expression choice:** To establish the correct parameters, we first deduce the eyebrow main action. We compute the average vertical value (y_{avg}) of the eyebrow for each frame and we compare it to the average vertical value computed when the eyebrow was in its neutral position. If the new average is greater, we will deduce that we have an *upward* expression; if it is smaller, we will deduce that we have a *downward* expression. The average value for the *neutral* expression arch is computed from the analysis of a frame previously labeled as neutral, on which the speaker did not move its eyebrows.

- **Parameter deduction:** Depending on the general expression we obtain the mathematical model parameters. They are deduced from comparing the thinned arch at the current frame against the arch of the eyebrow in a neutral position.

Parameter expressions:

– UPWARDS:

$$Ff_x \approx \Delta x[0] \cdot e^{(x_n[0]/\alpha)} \approx \Delta x[0] \quad \textbf{(6a)}$$

$$Ff_y' \approx \frac{\Delta y[N] - \Delta y[0]}{(e^{(-x_n[N]/\alpha)} - 1)} \quad \textbf{(6b)}$$

$$Ff_y \approx \Delta y[0] - Ff_y' \quad \textbf{(6c)}$$

– DOWNWARDS:

$$Fcs_x \approx -\Delta x[0] \quad \textbf{(7a)}$$

$$Foo_x \approx \frac{\Delta x[N] + Fcs_x}{\beta - x_n[N]} \quad \textbf{(7b)}$$

$$Fcs_y \approx \frac{-\Delta y[0] - \Delta y[N]}{2} \quad \textbf{(7c)}$$

$$Foo_y \approx \frac{-Fcs_y - \Delta y[\max]}{\beta^2} \quad \textbf{(7d)}$$

4 Verifying the Analysis Technique over Video Sequences

To our knowledge, it does not exist a database of face images that completely suits the needs of our tests. Nevertheless, we have tried to test our procedure over more than one speaker; and specifically, we show the results from the analysis of three individuals of different eyebrow characteristics taped under uncontrolled lighting conditions.

To test the correct behavior of the model and its application for eyebrow motion analysis, we applied the binarization-thinning technique over the left eye on the frames of several video sequences. Then, we deduced the model parameters contrasting the frame arch against the *neutral position arch*. To verify that the obtained parameters actually correspond to the eyebrow behavior, we have plotted the thinning results of each frame together with the obtained arch from applying the model over the *neutral position arch*, this resulting arch is the *modeled arch*. Figure 8 shows this arch comparison for the frames presented in the sequence of Figure 6; they also include the *neutral position arch*.

The best way to evaluate the performance of our techniques is to visually compare arch results; unfortunately, this procedure is not suitable to be applied over a large amount of data. To interpret the performance correctness of our approach, we have defined two different measurements:

(i) a *pseudo-area*:

$$\tilde{a} = \sum_i \left| y_k[i] - y_m[i] \right| \quad \textbf{(8)}$$

which can be understood as the area contained in-between arch k and m and it denotes the shape similarity between them; the closer $\tilde{a}$ is to 0 the more alike they are. We apply this measurement to check if the eye shape modeled by the extracted motion parameters follows the expected eyebrow obtained from the action;

(ii) 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 if the analysis procedure was not even able to detect the right eyebrow general action: up/down and it also provides information on the shape behavior of the *modeled arch*, by studying the sign of the measurement that gives a vertical location estimation.

We consider that the algorithm has worked right if the *pseudo-area* comparison shows that the *modeled arch* is closer to the *current frame arch* than the *neutral position arch*. Completely understanding the performance of this technique would also imply generating the synthetic eyebrow expression over a 3D head model of the speaker for a more detailed visual inspection. We will be able to perform this kind of tests when the pose coupling adaptation of the algorithm will be done and the tests will be performed on the complete analysis-synthesis of pose and expression system.

<u>Test conclusions</u>:

Results show that this analysis technique positively deduces the eyebrow behavior. We are able to analyze video images and extract the few needed parameters to understand and to later synthesize eyebrow motion.

From the visual inspection of our results we conclude that errors come more from the image processing performance of the analysis than from the motion model used. Correct binarization and later thinning are critical to obtain accurate motion parameters. Figure 9 plots the measurement results of three different tests. 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 low quality video input performance drops to around 50%. We must point out that the worst estimation usually happens for low expression movements, where the inaccuracy of the situation of the analysis area (the speaker may slightly move) is large enough to mislead the average results. In this case, like the *average difference* measurement shows, we may interpret an *up* movement as being *down* or vice versa.

Looking at Figure 7-b we realize how important the correct and precise definition of the eyebrow analysis area is. The graph plots the results of one analyzed sequence along with the neutral analyzed frame of another sequence where head location and size were not exactly the same. Motion not due to the eyebrow expression but to the overall head pose leads to mistaken results. Our tests have been performed accepting that the head pose on the video sequence is known and frontal. This is an unrealistic assumption if practical use of the algorithms is desired. Section 5 briefly explains which will be our next steps to adapt our algorithm to any given pose and to be able to use it within the telecom context exposed in the introduction.

5 Concluding Remarks and Future Work

The eyebrow motion analysis technique herein proposed has proved to give positive results. The animation parameters obtained represent eyebrow behavior and can be utilized to generate 3D-motion synthesis. To do so, these parameters need to be interpreted by some animation system and may have to be adapted to face animation parameter “semantics” of the animation engine. We have started to develop the adaptation procedure that allows us to generate the equivalent MPEG-4 FAP (face animation parameters).

The algorithms presented show two major weaknesses. First, we need to define one *neutral position frame* out of the video sequence from which we start the analysis comparisons. Second, all measurements are taken assuming that the pose remains unchanged along the time. In communication applications the analyzed face will freely move and such restriction will disappear.

In our complete analysis framework we have developed a head-tracking algorithm based on Kalman Filtering to predict the pose of the person. We also utilize a realistic 3D model of the speaker (clone) to generate an analysis by synthesis cooperation. This framework allows us to correctly track the region of interest of the features on each frame and to adapt our ‘near-to-front’ image analysis algorithms to any given pose after having verified their correct performance under the pose that gives the most expression information. We adapt these algorithms by redefining them on 3D using data extracted from the clones. We have already done this adaptation to the eye feature image analysis we had developed for a frontal position [7, 8] and we are currently developing and testing the adaptation procedure over the eyebrow analysis technique.

The use of highly realistic 3D head models will also allow us to get all the data taken for the *neutral position frame* from the analysis of synthesized images of the clone. The binarization process we have presented is suitable for white skin people with average eyebrows, but it may not work for people with very dark skin or almost unappreciable eyebrows. An ‘a priori’ color study of the speaker’s skin as well as the choice of the most suitable color space to perform the binarization skin/eyebrow could let us extend the complete procedure to any given case.

References

1. DeCarlo, D., Metaxas, D.: Optical Flow Constraints on Deformable Models with Applications to Face Tracking. In *International Journal of Computer Vision* (July 2000) 38(2) 99-127
2. Eisert, P., Girod, B.: Analyzing Facial Expressions for Virtual Conferencing. In *IEEE Computer Graphics & Applications* (September 1998) 70-78
3. Valente, S., Dugelay, J.-L.: Face Tracking and Realistic Animations for Telecommunicant Clones. In *IEEE Multimedia Magazine* (February 2000) 34-43
4. Dugelay, J.-L., Andrés del Valle, A. C.: Analysis-Synthesis Cooperation for MPEG-4 Realistic Clone Animation. In Proceedings of the *Euroimage ICAV3D* (June 2001)
5. Tian, Y., Kanade, T., Cohn, J.: Recognizing Lower Face Action Units for Facial Expression Analysis. In Proceedings of the *4th IEEE International Conference On Automatic and Gesture Recognition* (March 2000) 484-490
6. Goto, T., Escher, M., Zanardi, C., Magnenat-Thalmann, N.: MPEG-4 Based Animation with Face Feature Tracking. In Proceedings of the *Eurographics Workshop on Animation and Simulation* (September 1999) Springer-Verlag 89-98
7. Andrés del Valle, A. C., Dugelay, J.-L.: Eye State Tracking for Face Cloning. In Proceedings of the *International Conference on Image Processing* (October 2001)
8. Andrés del Valle, A. C., Dugelay, J.-L.: Facial Expression Analysis Robust to 3D Head Pose Motion. In Proceedings of the *International Conference on Multimedia and Expo* (August 2002)

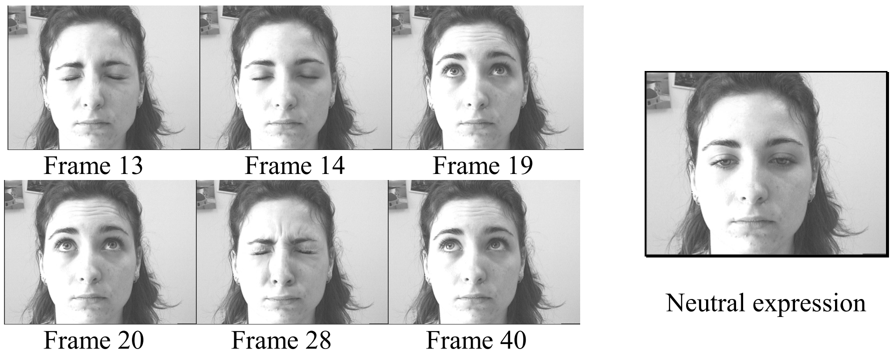

Fig. 6. Our tests were performed over video sequences where the lighting over the face was not uniform. No environmental conditions were known besides the exact location of the ROI including the eyebrow feature, which remained unchanged through the sequence. Here we present the frames analyzed to obtained the results presented in Figure 8

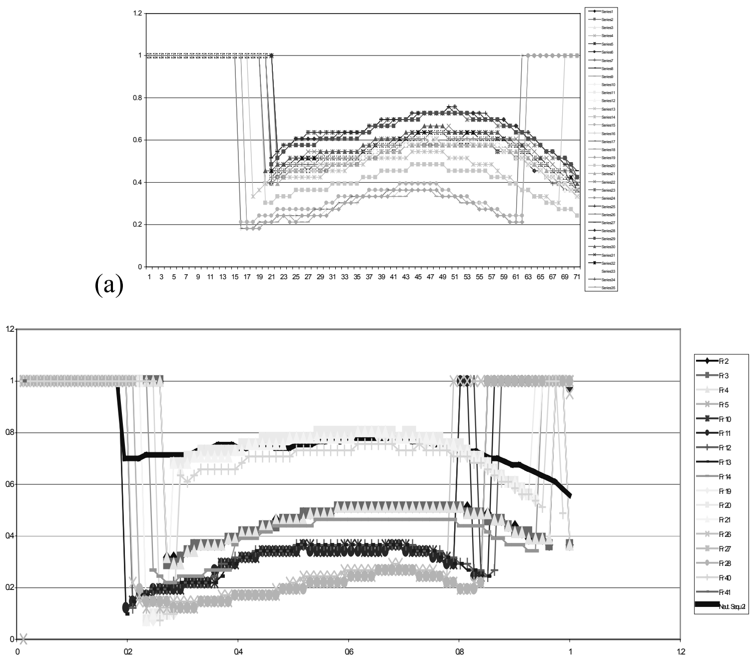

Fig. 7. Correct binarization and thinning clearly gives the data from which to extract the model parameters. Graph (b) plots the mixed results from the analysis of two different video sequences. **Neut. Seq.2** is the analysis of a frame where the eyebrow was relaxed taken from a sequence different from the **Fr sequence**. This comparison simulates what would happen if the pose of the speaker changed during the analysis. The pose motion would cause the movement of the eyebrow but the algorithm would interpret it as a local eyebrow expression (being *upwards* when in reality it is neutral) We must control the pose of the user to completely exploit the algorithm in practical applications

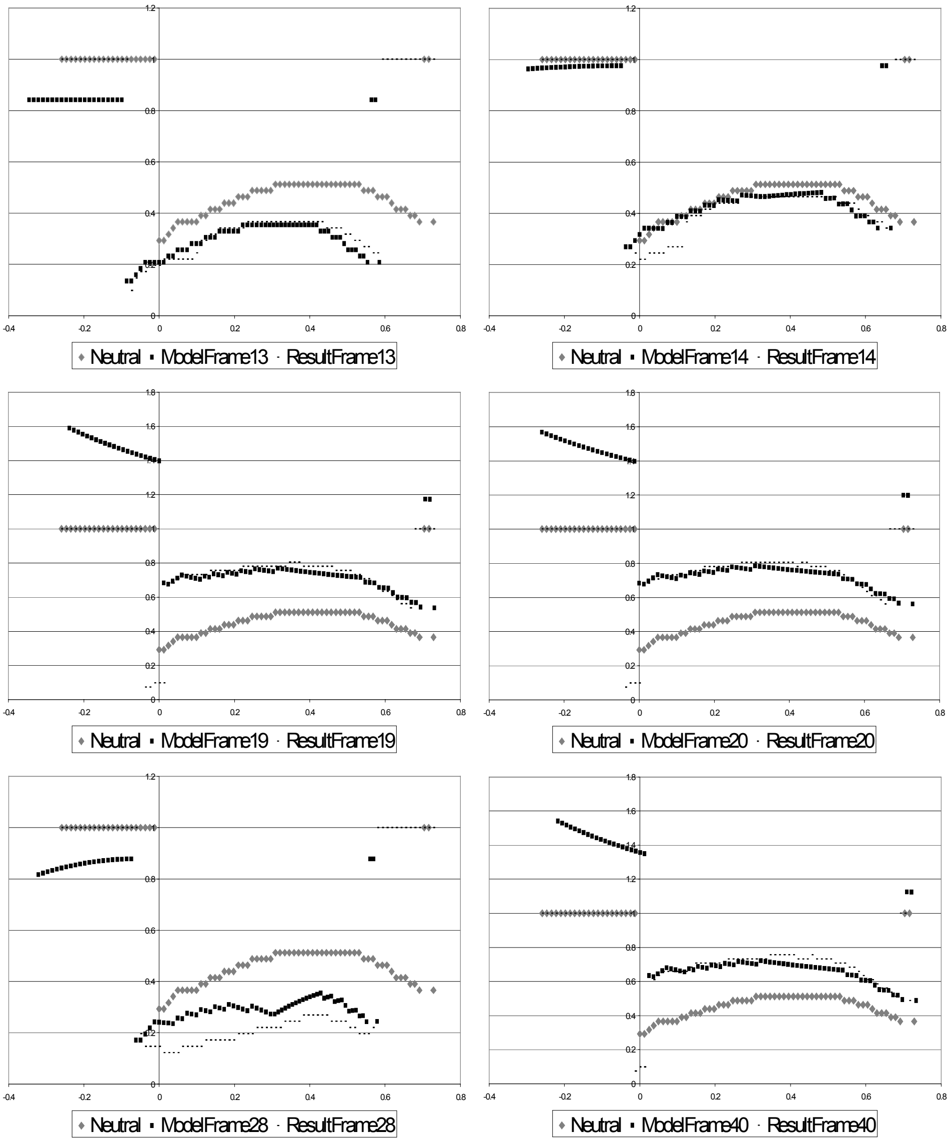

Fig. 8. The anatomic-mathematical motion model nicely represents the eyebrow deformation. We see on frame 28 how the strange thinning result obtained at the beginning of the arch, probably due to the eyebrow-eye blending during binarization, worsens the algorithm accuracy. Although the obtained parameters still correctly interpret the general downward movement, showing fair robustness, they are no longer able to express the exact motion intensity

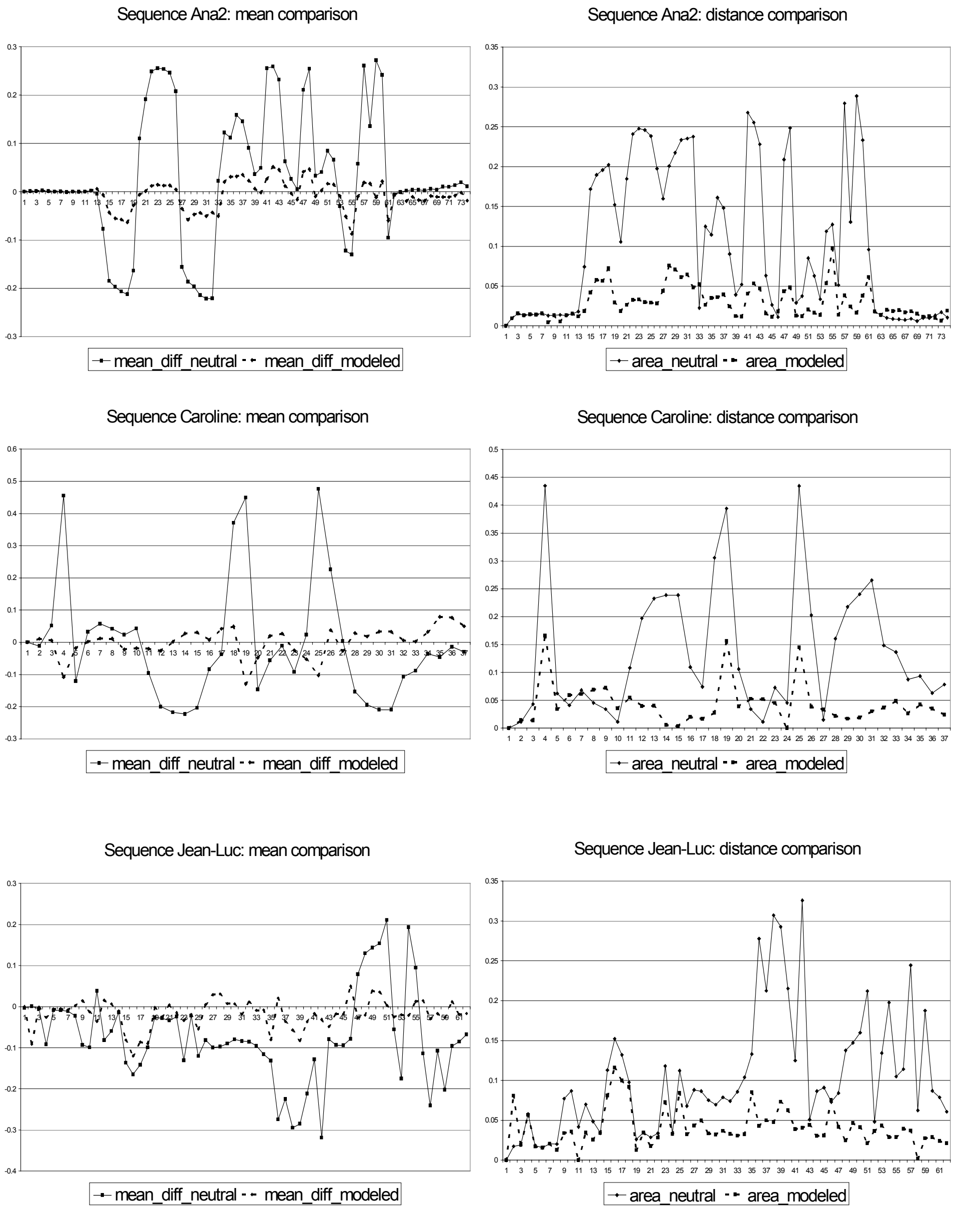

Fig. 9. These plotted results from three different sequences: Ana2, Caroline and Jean-Luc illustrate the analysis behavior of the algorithm under different conditions. The algorithm proves to detect the right movement (the mean difference decreases) and to estimate the motion parameters correctly (the area decreases). We observe the best behavior for extreme eyebrow expressions. Ana2 sequence success rate: 90.71%, Caroline sequence success rate: 78.38% and Jean-Luc sequence success rate: 82.26%