CAR-TR-756 DACA76-92-C-0009
CS-TR-3401 January 1995

Tracking and Recognizing Facial Expressions in Image Sequences, using Local Parameterized Models of Image Motion

Michael J. Black
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

Yaser Yacoob
Computer Vision Laboratory
Center for Automation Research
University of Maryland
College Park, MD 20742-3275

Abstract

We explore the use of local parametrized models of image motion for recovering and recognizing the non-rigid and articulated motion of human faces. Parametric flow models (for example, affine) are popular for estimating motion in rigid scenes. We observe that within local regions in space and time, such models not only accurately model non-rigid facial motions but also provide a concise description of the motion in terms of a small number of parameters. These parameters are intuitively related to the motions of facial features during facial expressions. We show how expressions such as anger, happiness, surprise, fear, disgust, and sadness can be recognized from the local parametric motions in the presence of significant head motion. Our motion tracking and expression recognition approach performed with high accuracy in extensive laboratory experiments involving 40 subjects as well as in television and movie sequences.

Keywords: Dynamic facial expression recognition, parametric models of optical flow, non-rigid motion.

The support of the Advanced Research Projects Agency (ARPA Order No. 8459) and the U.S. Army Topographic Engineering Center under Contract DACA76-92-C-0009 is gratefully acknowledged, as is the help of Sandy German in preparing this paper.

1 Introduction

The recognition of facial expressions in image sequences involving significant head motion is a challenging problem with many applications in human-computer interaction. Yet, while the coincidence of head and facial feature motion is prevalent in human behavior, it has so far attracted very little attention as a motion estimation problem. Previous work has typically focused on only one part of the problem or the other: either on rigid head tracking [2, 3] with no facial expression, or on expression recognition with no motion at all [27] or with a roughly stationary head [22, 25].

Here we propose a simple model of rigid and non-rigid facial motion using a collection of local parametric models. The image motion of the face, mouth, eyebrows, and eyes are modeled using image flow models with only a handful of parameters. The motions of these regions are estimated over an image sequence using a robust regression scheme [6] which makes the recovered motion parameters stable under adverse conditions such as motion blur, blooming, loss of focus, etc. These recovered image motion parameters correspond simply and intuitively to various facial expressions. We illustrate how these parameters can be used to recognize facial expressions even in situations where the motion of the head is large.

Models used in recognizing facial expressions vary in the amount of geometric information about head shape and motion they contain. At one extreme are approaches which employ physically-based models of heads including skin and musculature [15, 22]. A slightly weaker model uses deformable templates to represent feature shapes in the image [27]. At the other extreme is the work of Yacoob and Davis [25] in which they recognize facial expressions using statistical properties of the optical flow with only very weak models of facial shape. In this paper we explore a middle ground between the template-based approaches and the optical flow-based approaches in which we represent rigid and deformable facial motions using piecewise parametric models of image motion. These models provide greater abstraction and robustness than the purely flow-based methods yet are weaker than models which incorporate detailed information about shape. While parametric flow models (for example affine) are popular for estimating motion in rigid scenes [5, 6, 8], their application to non-rigid motion is unconventional. However, within local regions in space and time, such models not only accurately model non-rigid facial motions but also provide a concise description of the motion in terms of a small number of parameters.

To model the rigid motion of a face in an image we make the simple assumption that the majority of the face can be modeled by a plane (see Figure 1). More complex models can be employed (for example an ellipsoid) but the planar model is particularly simple: the image motion of a plane can be described by eight parameters. A face is neither planar nor strictly rigid but, when robust estimation techniques are employed, the simple planar model can be used in situations where there are outliers due to non-planarity or non-rigidity. The planar model is sufficient for recovering qualitative information about the motion of the head. More sophisticated models could be used if accurate information about the 3D motion of the head is required.

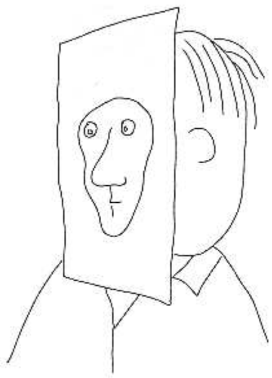


Figure 1: The rigid component of the face motion is modeled as the motion of a plane.

The image motions of the facial features (eyes, mouth and eyebrows) are modeled relative to the head motion using different parametric models. For the eyes a simple affine model is used. For the brows and mouth an affine model is augmented with an additional curvature parameter to account for the arching of the brows and curvature of the mouth during smiling. Additionally, the nose region can be tracked, if desired, using an affine model. In this paper we do not address the problem of initially locating the various facial features; this topic has been addressed in [9, 24, 28]. Notice that while the motion of the entire face can be quite complicated, when it is broken down into parts, the motion of each part can be modeled very simply.

The approach is summarized as follows (and is illustrated in Figure 2). Given the location of the face, eyes, brows, and mouth, estimate the rigid motion of the face region (excluding the deformable features) between two frames using a planar motion model. This estimation is performed using a robust statistical approach to cope with violations of the rigid plane assumption. The motion of the face is used to register the images via warping and then the relative motion of the feature regions is estimated in the coordinate frame of the face using the same robust estimation procedure. The motion estimates of the face and features are used to predict their locations in the next frame and the process is repeated. The estimated motion parameters provide a simple abstraction of the underlying facial motions and can be used to classify the type of rigid head motion and the facial expression.

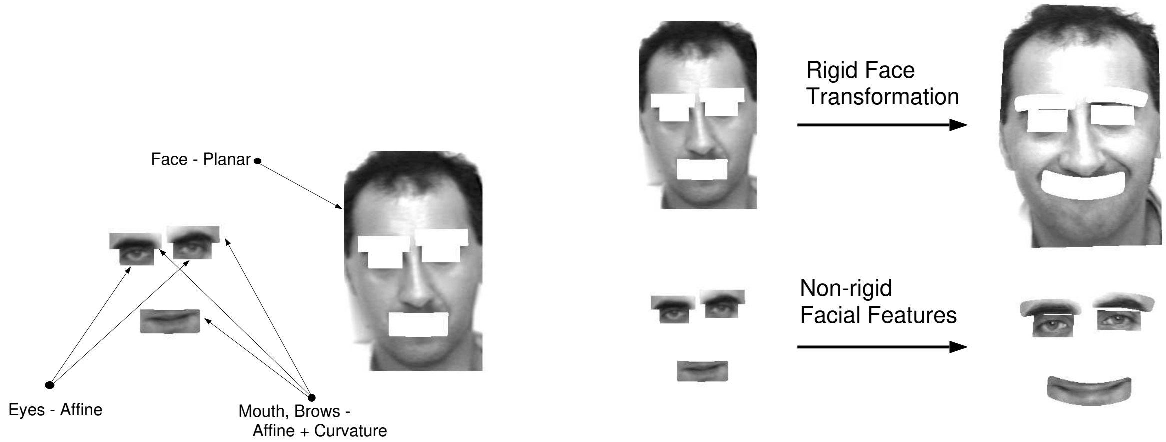


Figure 2: Illustration showing the parametric motion models employed and and example of a face undergoing a looming motion while smiling.

In the next section we review previous work on facial feature tracking and expression recognition. Section 3 presents the parametric motion models used for the various facial features while Section 4 addresses the robust estimation scheme used to recover the motion parameters. These parameters are related to facial expressions in Section 5 and then Section 6 discusses the results of tracking experiments with various expressions and rigid head motions. Section 7 discusses our mid- and high-level representations of facial motions as well as the recognition process. The results of extensive testing of the recognition method are presented in Section 8.

2 Previous Work: Human facial tracking

Head and facial feature tracking have been viewed as a spatio-temporal problem where the initial locations of the head and features are assumed to be known. Head tracking involves tracking the motion of a rigid object performing rotations and translations while facial feature tracking involves tracking diverse non-rigid deformations that are limited by the anatomy of the head. There are two basic types of feature tracking: *feature boundary* and *feature region* tracking. Feature boundary tracking attempts to track and accurately delineate the shapes of the facial features—e.g., to track the contours of the lips and mouth opening [18, 22]. Feature region tracking, on the other hand, addresses the simpler problem of tracking a region, such as a bounding box, surrounding the facial feature, paying little if any attention to the detailed shape of the feature. In this paper we focus on the latter tracking approach, and show that it allows us to capture and describe several aspects of the rich repertoire of facial expressions.

Facial features are subject to one or more of the following types of motions: rigid, articulated, and deformable motions. The rigid motion is due to the head's translation or rotation. The articulated motion includes the motion of the lower jaw during speech and several facial expressions (e.g., the 'surprise' expression). Typical deformable motions are due to muscle contractions and expansions that accompany speech and facial expressions.

An approach to rigid head tracking and motion estimation by tracking points with high Hessian values was proposed in [3]. Several such points are tracked over the head and the 3D motion parameters of the head are recovered by solving an over-constrained set of motion equations.

Essa *et al.* [14] proposed a 3D model-based approach for tracking facial features. They assumed that a mesh was placed on the face and used the optical flow field to displace the mesh vertices and recover the locations of points on the face during the facial deformation.

Li *et al.* [20] proposed a model-based approach that assumes that a 3D mesh has been placed on the face in the image, and that the depths of points on the face have been recovered. They proposed algorithms for recovering the rigid and non-rigid motions of the face from the sequence of images, and reapplying these motions to create an approximation to the initial sequence. Their model-based approach employed knowledge about the anatomy of the face to constrain the estimation of the non-rigid facial motion.

The facial feature tracking experiments reported in [21, 25] were based on analysis of the magnitudes of the gradients of the intensity image and the optical flow fields of the image sequence. The changes in these values between consecutive images provided clues to the spatial change of each facial feature. This approach dealt well with articulated and deformable motions, but was able to accommodate only limited rigid motion.

The work reported in [22] assumes that eleven principal contours are initially located (in practice manually) on the face. These contours are tracked throughout the sequence by applying an image force field that is computed from the gradient of the intensity image. In addition to assuming a frontal view, it was assumed that the projection is orthographic, and some facial make-up was used.

Our objective is to develop a passive system that is able to track a human head and primary facial features in a dynamic environment and provide a rich description of the observed motions. The description we seek will allow separation between different classes of motion so that reasoning about facial expression and head gesture is feasible. In this respect, our goal requires far more accurate motion analysis than the more traditional goal of resynthesis of the facial actions (which requires a sequence visually similar to the source sequence).

In contrast to the 3D based estimation of some of these motions as in [14, 15, 20, 22] we will focus on 2D recovery of the components of the motions.

3 Models of Image Motion

The estimation of image motion requires the integration of information over some neighborhood in the image under some assumptions about the variation in the motion. Parameterized models of image motion make explicit the assumptions about the motion and typically assume that the image flow can be represented by a low-order polynomial [5]. Within small image regions the following affine model of image motion is often sufficient [19]:

$$u(x,y) = a_0 + a_1 x + a_2 y \qquad (1)$$

$$v(x, y) = a_3 + a_4 x + a_5 y \tag{2}$$

where the a_i are constants and where $u(x, y)$ and $v(x, y)$ are the horizontal and vertical components of the flow at the image point $\mathbf{x} = (x, y)$.

The parameters a_i have simple interpretations in terms of image motion. For example, a_0 and a_3 represent horizontal and vertical translation respectively. Additionally, we can express *divergence* (isotropic expansion), *curl* (rotation about the viewing direction), and *deformation* (squashing or stretching) as combinations of the a_i [10, 19]:

$$\text{divergence} = a_1 + a_5 = (u_x + v_y), \tag{3}$$

$$\text{curl} = -a_2 + a_4 = -(u_y - v_x), \tag{4}$$

$$\text{deformation} = a_1 - a_5 = (u_x - v_y) \tag{5}$$

where the subscripts x and y indicate partial derivatives of the image velocity. Divergence, curl and the magnitude of the deformation have the convenient property of being invariant to rotations of the image coordinate frame [10]. Translation, along with divergence, curl, and deformation, will prove to be useful for describing facial expressions and are illustrated in Figure 3. For example, when the motion of the eye regions is modeled as being affine, eye blinking can be detected as rapid deformation, divergence, and vertical translation in the eye region.

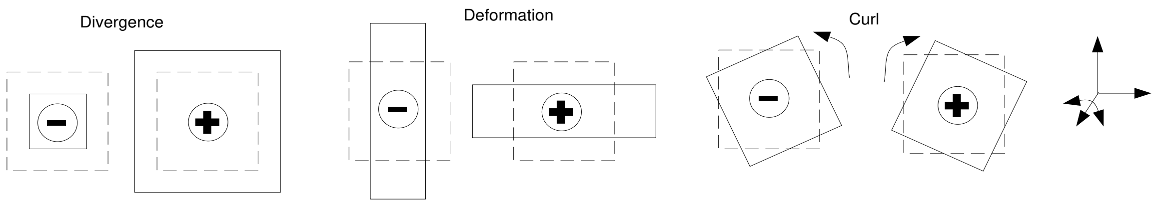


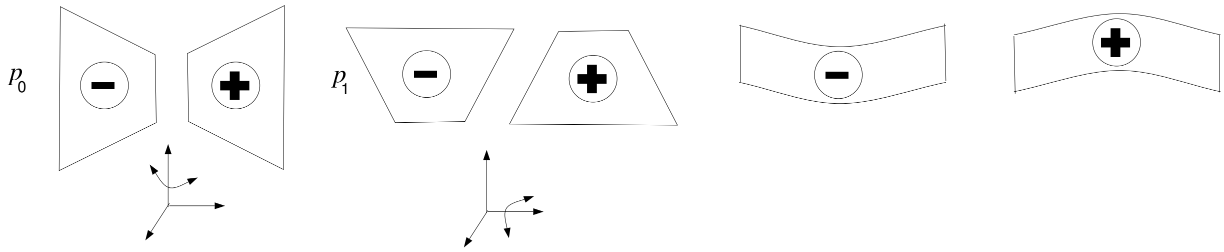
Figure 3: The figure illustrates the motion captured by the various parameters used to represent the motion of the regions. The solid lines indicate the deformed image region and the "–" and "+" indicate the sign of the quantity.

The affine model is not sufficient to capture the image motion of a human face when it occupies a significant portion of the field of view. A more appropriate model (which is still a gross approximation to face shape) would assume that the face is a plane viewed under perspective projection. It is well known that the image motion of a rigid planar region of the scene can be described by the following eight-parameter model [1, 23]:

$$u(x, y) = a_0 + a_1 x + a_2 y + p_0 x^2 + p_1 xy, \tag{6}$$

$$v(x, y) = a_3 + a_4 x + a_5 y + p_0 xy + p_1 y^2, \tag{7}$$

where we have added two new terms p_0 and p_1 to the affine model. These parameters roughly represent "yaw" and "pitch" respectively and are illustrated in Figure 4. While we have experimented with more complex models of rigid face motion we have found that the planar assumption is both simple and expressive enough to robustly represent qualitative rigid facial motions in a variety of situations.



Planar parameters

Curvature parameter c

Figure 4: Additional parameters for planar motion and curvature.

Non-rigid motions of facial features such as the eyebrows and mouth, however, are not well captured by the rigid affine or planar models. Deformable models such as snakes provide good tracking of these regions [22] but their distributed nature does not admit simple, intuitive, characterizations of the motions, such as we saw above, but rather necessitates additional analysis to extract meaningful descriptions of the motion. An alternative would be to use parameterized curve models for tracking. Deformable templates [27], on the other hand, encode information about shape but not about motion. We wish to stay within the paradigm of using parametric models of image motion and so we augment the affine model to account for the primary form of curvature seen in mouths and eyebrows. We add a new parameter c to the affine model:

$$u(x, y) = a_0 + a_1 x + a_2 y \tag{8}$$

$$v(x, y) = a_3 + a_4 x + a_5 y + cx^2 \tag{9}$$

where c encodes curvature and is illustrated in Figure 4. As our experiments will demonstrate, this seven-parameter model captures the essential image motion of the mouth and eyebrows.

Unfortunately this new curvature parameter is not invariant to head rotations. The curvature of the mouth and eyebrows should roughly be oriented with the principal axis of the face. To estimate the curvature with respect to the coordinate frame of the face we compute the orientation of the principal axis of the face and transform the images and features into the coordinate frame of the image plane. We then estimate the curvature and transform the features back into the coordinate frame of the face for purposes of tracking.

4 Robust Regression

For convenience of notation we define

$$\begin{aligned}
\mathbf{X}(\mathbf{x}) &= \begin{bmatrix} 1 & x & y & 0 & 0 & 0 & x^2 & xy & 0 \\ 0 & 0 & 0 & 1 & x & y & xy & y^2 & x^2 \end{bmatrix}, & (10)\\
\mathbf{A} &= \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & 0 & 0 & 0 \end{bmatrix}^T, & (11)\\
\mathbf{P} &= \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & p_0 & p_1 & 0 \end{bmatrix}^T, & (12)\\
\mathbf{C} &= \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & 0 & 0 & c \end{bmatrix}^T, & (13)
\end{aligned}$$

such that $\mathbf{u}(\mathbf{x};\mathbf{A}) = \mathbf{X}(\mathbf{x})\mathbf{A}$, $\mathbf{u}(\mathbf{x};\mathbf{P}) = \mathbf{X}(\mathbf{x})\mathbf{P}$, and $\mathbf{u}(\mathbf{x};\mathbf{C}) = \mathbf{X}(\mathbf{x})\mathbf{C}$ represent the affine, planar, and affine+curvature flow models described in the previous section.

Let f be the set of image points corresponding to the face region (excluding the non-rigid features), and P_f the planar motion parameters of these points. The brightness constancy assumption for the face states that

$$I(\mathbf{x}, t) = I(\mathbf{x} - \mathbf{X}(\mathbf{x})\mathbf{P}_f, t+1), \quad \forall \mathbf{x} \in f, \tag{14}$$

where I is the image brightness function and t represents time. Taking the Taylor series expansion of the right hand side, simplifying, and dropping terms above first order gives

$$\nabla I(\mathbf{X}(\mathbf{x})\mathbf{P}_f) + I_t = 0, \quad \forall \mathbf{x} \in f, \tag{15}$$

where $\nabla I = [I_x, I_y]$ and the subscripts indicate partial derivatives of image brightness with respect to the spatial dimensions and time.

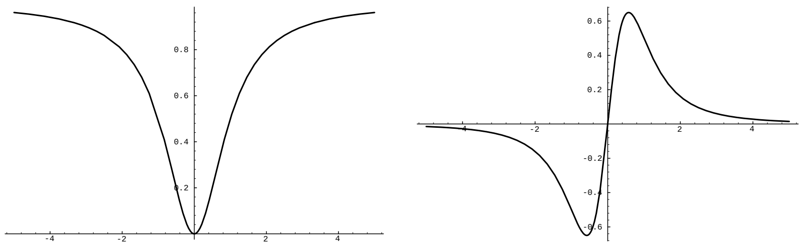
To estimate the parameters $\mathbf{P}_f$ we minimize

$$\sum_{\mathbf{X}\in f} \rho(\nabla I(\mathbf{X}(\mathbf{x})\mathbf{P}_f) + I_t, \sigma), \tag{16}$$

for some error norm ρ where σ is a scale parameter. Since the face is neither a plane nor rigid it is important to take ρ to be a robust error norm which can cope with some percentage of gross errors or "outliers" [17]. For the experiments in this paper we take ρ to be

$$\rho(x, \sigma) = \frac{x^2}{\sigma + x^2} \tag{17}$$

which is the robust error norm used by Geman and McClure [16]; it is shown in Figure 5. As the magnitudes of the residuals $\nabla I(\mathbf{X}(\mathbf{x})\mathbf{P}_f) + I_t$ grow beyond a certain point their influence on the solution begins to decrease and the value of $\rho(\cdot)$ approaches a constant. The function $\psi(x, \sigma)$ shown in Figure 5 is the derivative of ρ and characterizes the influence of the residuals [17]. The value σ is a scale parameter that effects the point at which the influence of outliers begins to decrease.



$\rho(x, \sigma)$ $\psi(x, \sigma)$

Figure 5: Geman-McClure robust error norm (ρ) and influence function (ψ).

Equation (16) is minimized using a simple gradient descent scheme with a continuation method that begins with a high value for σ and lowers it during the minimization (see [6, 7] for details). The effect of this procedure is that initially no data are rejected as outliers; then the influence of outliers is gradually reduced. To cope with large motions a coarse-to-fine strategy is used in which the motion is first estimated at a coarse level; then, at the next finer level, the image at time $t + 1$ is warped towards the image at time t using the current motion estimate. The motion parameters are refined at this level and the process continues until the finest level.

Once the face motion is estimated it is used to register the image at time $t+1$ with the image at time t by warping the image at $t+1$ back towards the image at t. Then the relative motions of the facial features are estimated using the registered images.

4.1 Tracking facial features

The motions of the face and facial features estimated between two frames are used to predict the locations of the features in the next frame. The face and the eyes are simple quadrilaterals which are represented by the image locations of their four corners. Since a line on a plane remains a line under the planar motion $\mathbf{P}_f$ these regions remain quadrilaterals although the locations of their four corners change. We update the location $\mathbf{x}$ of each of the four corners of the face and eyes by applying the planar motion to get $\mathbf{X}(\mathbf{x})\mathbf{P}_f + \mathbf{x}$. Then the relative motion of the eyes' locations is accounted for and the corners become $(\mathbf{X}(\mathbf{x})\mathbf{P})\mathbf{A}_{le} + \mathbf{x}$ and $(\mathbf{X}(\mathbf{x})\mathbf{P})\mathbf{A}_{re} + \mathbf{x}$ where le and re stand for the motions of the left and right eyes respectively. In updating the eye region we do not use the full affine model since when the eye blinks this would cause the tracked region to deform to the point where the eye region could no longer be tracked. Instead only the horizontal and vertical translation of the eye region are used to update its location relative to the face motion.

The curvature of the mouth and brows implies that simple updating of the corners is not sufficient for tracking. In our current implementation we use image masks to represent the regions of the image corresponding to the brows and the mouth. These masks are updated by warping them first by the planar face motion $\mathbf{P}_f$ and then by the motion of the individual features $\mathbf{C}_m$, $\mathbf{C}_{lb}$ and $\mathbf{C}_{rb}$ which correspond to the mouth and the left and right brows respectively. This simple updating scheme works well in practice.

To reduce noise in the parameters we use a simple temporal filter. Let $\mathbf{P}_f^+$ be the filtered parameters of the face and $\mathbf{P}_f$ be the current estimate of the face parameters; then we update $\mathbf{P}_f^+$ as follows:

$$\mathbf{P}_f^+ = \frac{1}{2}(\mathbf{P}_f^+ + \mathbf{P}_f).$$

Exactly the same treatment is applied to the relative facial feature motions and these smoothed values are used for expression recognition. A more sophisticated Kalman filter could be used as in [3] but this averaging scheme works well in our experiments and has the property of weighting current estimates more heavily than previous ones. This is an appropriate model for facial expressions which are typically of short duration.

5 Expression recognition

Research in psychology has indicated that at least six emotions are universally associated with distinct facial expressions [12]. Several other emotions, and many combinations of emotions, have been studied but remain unconfirmed as universally distinguishable. The six principal emotions are: happiness, sadness, surprise, fear, anger, and disgust.

Most psychological research on facial expressions has been conducted on "mug-shot" pictures that capture the subject's expression at its peak [26]. These pictures allow one to detect the presence of static cues (such as wrinkles) as well as the positions and shapes of the facial features. Few studies have directly investigated the influence of the motion and deformation of facial features on the interpretation of facial expressions. Bassili [4] suggested that motion in the image of a face would allow emotions to be identified even with minimal information about the spatial arrangement of features. The subjects of his experiments viewed image sequences in which only white dots on the dark surface of the face displaying the emotion are visible. The reported results indicate that facial expressions were more accurately recognized from dynamic images than from a single static image. Whereas all expressions were recognized at above chance levels (i.e., above 16%) in dynamic images, only happiness and sadness were recognized at above chance levels in static images.

In developing our approach to expression recognition we relied on the psychological sources best represented for static images by the work of Ekman and Friesen [13], and for image sequences by the work of Bassili [4], as well as on the recent mid- and high-level spatio-temporal representations proposed by Yacoob and Davis [25].

The observations found in [13] are summarized in Table 1. They describe observations made on single pictures of faces, though it should be noticed that occasionally implicit temporal descriptions are used.

Figure 6 summarizes the observations of Bassili [4] on motion-based cues for facial expres-

Table 1: The cues for facial expressions as suggested by Ekman and Friesen.

Emotion	Observed facial cues
Surprise	brows raised (curved and high) skin below brow stretched horizontal wrinkles across forehead eyelids opened and more of the white of the eye is visible jaw drops open without tension or stretching of the mouth
Fear	brows raised and drawn together forehead wrinkles drawn to the center upper eyelid is raised and lower eyelid is drawn up mouth is open lips are slightly tense or stretched and drawn back
Disgust	upper lip is raised lower lip is raised and pushed up to upper lip or is lowered nose is wrinkled cheeks are raised lines below the lower lid, lid is pushed up but not tense brows are lowered, lowering the upper lid
Anger	brows lowered and drawn together vertical lines appear between brows lower lid is tensed and may or may not be raised upper lid is tense and may or may not be lowered due to brows' action eyes have a hard stare and may have a bulging appearance lips are either pressed firmly together with corners straight or down or open, tensed in a squarish shape nostrils may be dilated (could occur in sadness too) unambiguous only if registered in all three facial areas
Happiness	corners of lips are drawn back and up mouth may or may not be parted with teeth exposed or not a wrinkle runs down from the nose to the outer edge beyond lip corners cheeks are raised lower eyelid shows wrinkles below it, and may be raised but not tense crow's-feet wrinkles go outward from the outer corners of the eyes
Sadness	inner corners of eyebrows are drawn up skin below the eyebrow is triangulated, with inner corner up upper lid inner corner is raised corners of the lips are drawn or lip is trembling

sions. Recall that the experiments of Bassili were intended to explore only the role of motion in facial expressions; therefore the face features, texture and complexion were unavailable to the subjects. As illustrated in Figure 6, Bassili identified principal facial motions that provide powerful cues to the subjects in recognizing facial expressions.

Following the temporal approach to recognition proposed in [25], we divided each facial

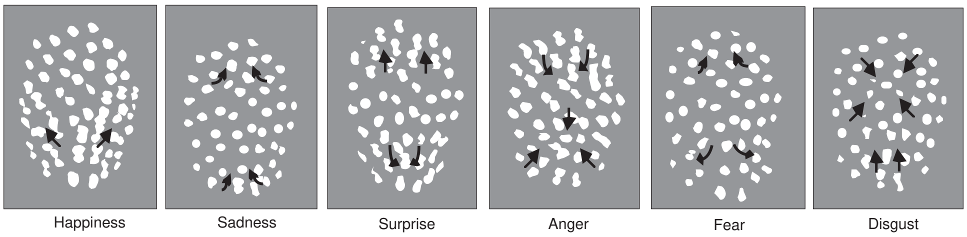


Figure 6: The cues for facial expression as suggested by Bassili.

expression into three temporal segments: its *beginning*, *apex* and *ending*. Figure 7 illustrates the temporal segments of a smile model. Notice that Figure 7 indicates that the change in parameter values might not occur at the same frames at either the beginnings or endings of actions, but it is required that a significant overlap be detectable to label a set of frames with a "beginning of a smile" label, while the motions must terminate before a frame is labeled as an "apex" or an "ending."

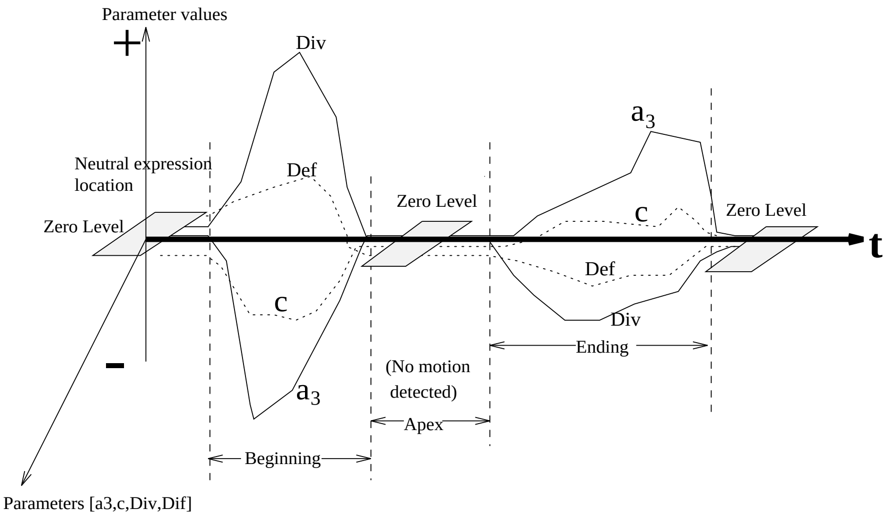


Figure 7: The temporal model of the "smile" expression.

The detailed development of the smile model is as follows. The upward-outward motion of the mouth corners results in a negative curvature of the mouth (i.e., the curvature parameter c is negative). The horizontal and overall vertical stretching are manifested by positive divergence (Div) and deformation (Def). Finally, some overall upward translation is caused by the raising of the lower and upper lips due to the stretching of the mouth (a_3 is negative). Reversal of these motion parameters is observed during the ending of the expression.

Table 2 shows the rules used to identify the onsets of the "beginning" and the "ending" of three facial expressions ('smile,' 'anger' and 'surprise'). These rules are applied to the computed parameters to create a complete temporal map describing the evolving facial expression.

Table 2: The parameter values used for classifying facial expressions (B=beginning, E=ending). +/− means the value is positive/negative, 0 means it should be small, and blank indicates "don't care").

Expr.	B/E	Feature	a_0	a_3	Div	Curl	Def	c
Anger	B	Mouth		−		0	+	−
		R. Brow	+	+		+	+	−
		L. Brow	−	+		−	+	−
		R. Eye	+		−		+	
		L. Eye	−		−		+	
Anger	E	Mouth		+		0	−	+
		R. Brow	−	−		−	−	+
		L. Brow	+	−		+	−	+
		R. Eye	−		+		−	
		L. Eye	+		+		−	
Happiness	B	Mouth		−			+	−
Happiness	E	Mouth		+			−	+
Surprise	B	Mouth		+	+	0	−	
		R. Brow	−	−		−		+
		L. Brow	+	−		+		+
		R. Eye	−	−	+		−	
		L. Eye	+	−	+		−	
Surprise	E	Mouth		−	−	0	+	
		R. Brow	+	+		+		−
		L. Brow	−	+		−		−
		R. Eye	+	+	−		+	
		L. Eye	−	+	−		+	

To simplify the illustration of our approach we consider in Table 2 only one class of the several forms of expressions for 'smile' and 'anger.' The parameter values for a 'smile' do not consider whether the smile is a genuine 'smile' or a posed one (more on this subject can be found in [13]). Similarly, we consider an 'anger' expression in which the mouth is tightened, although some expressions of 'anger' that include 'rage' and 'aggressiveness' may show the mouth slightly opened, and tensed in a squarish shape (see Table 1). These and other behaviors will be modeled in our future work.

To minimize the effect of noise, the parameters used in Table 2 are typically thresholded so that small values are not considered.

6 Motion Estimation Results

The experiments in this section illustrate the rigid and non-rigid motion tracking while highlighting the information contained in the motion parameters that can be used for expression recognition. The first four experiments illustrate the relationship between the motion parameters and specific facial expressions (Happiness, Anger, Surprise, and Blinking) in the simple situation where the head is relatively stable. These experiments are followed by longer experiments on sequences of 100 or more images in which there is rigid head motion combined with the non-rigid motions of facial expressions. The reader is referred to Figures 3 and 4 in Section 3 to aid in understanding the parameters plotted in this section.

All parameter settings used in the motion estimation algorithm were exactly the same in all the experiments. In particular, for each pair of images, a Gaussian pyramid was constructed and 15 iterations of gradient descent were used at each of the four levels in the coarse-to-fine strategy. The motion at one level is used to register the images before refining the estimate at the next finer level. The value of σ began at $15.0\sqrt{3}$, was lowered by a factor of 0.95 after each iteration, and was reset at each level of the pyramid. The initial regions for the head and the features were selected by hand and were automatically tracked thereafter.

6.1 Smile experiment

The facial features tracked through the 'smile' image sequence are shown in Figure 8. The figure shows the image regions corresponding to each of the features at a particular frame. As the 'smile' begins the curvature of the mouth is clearly visible; then this curvature is followed by an elongation of the mouth.

Figure 9 plots some of the parameters of the mouth motion. The left graph shows the horizontal (solid line) and vertical (dashed line) translation of the mouth. The negative vertical translation indicates that the mouth rises relative to the face during the initial phase of the 'smile' expression. The middle graph plots curvature of the mouth (parameter

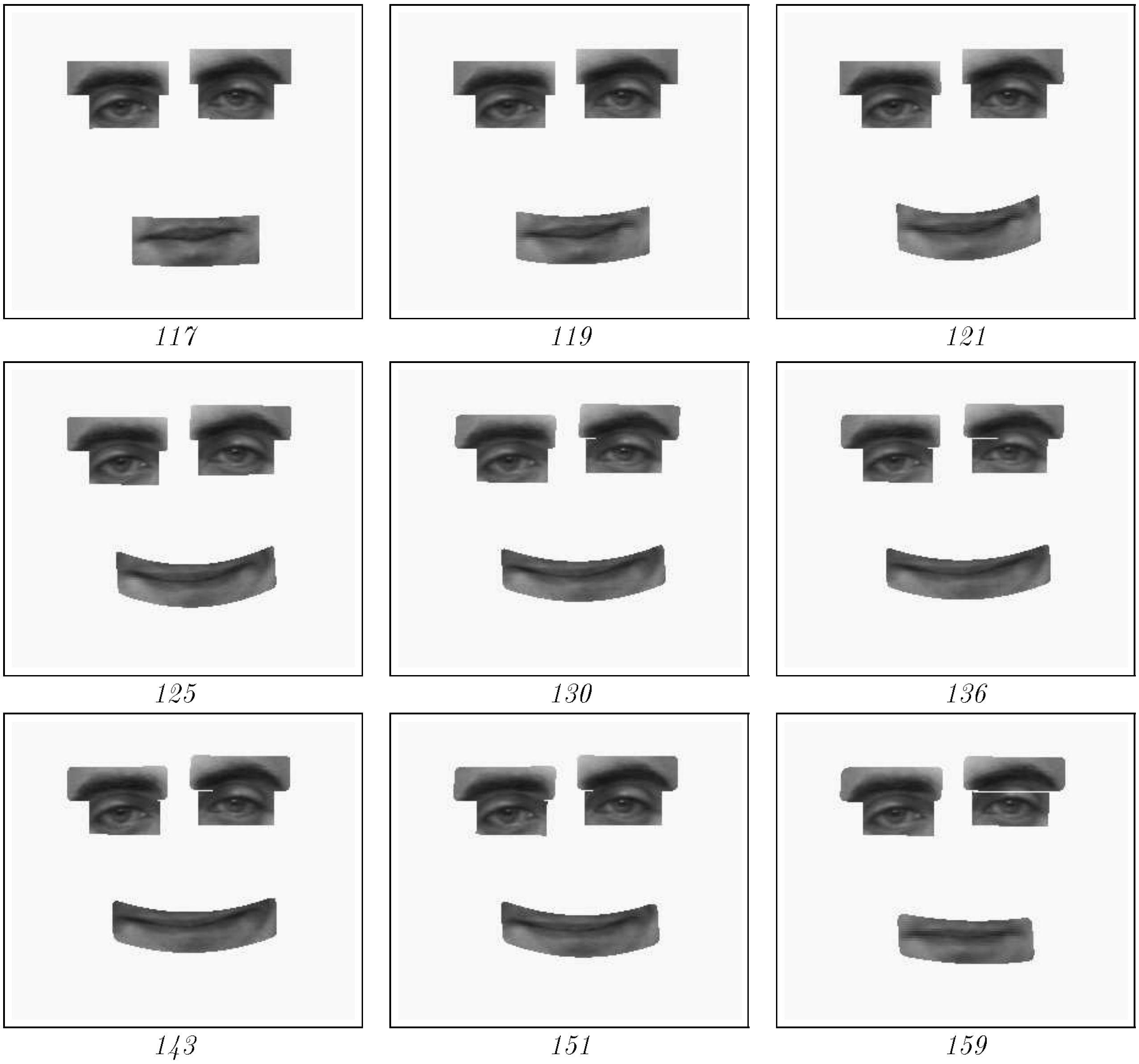


Figure 8: Smile experiment: facial expression tracking.

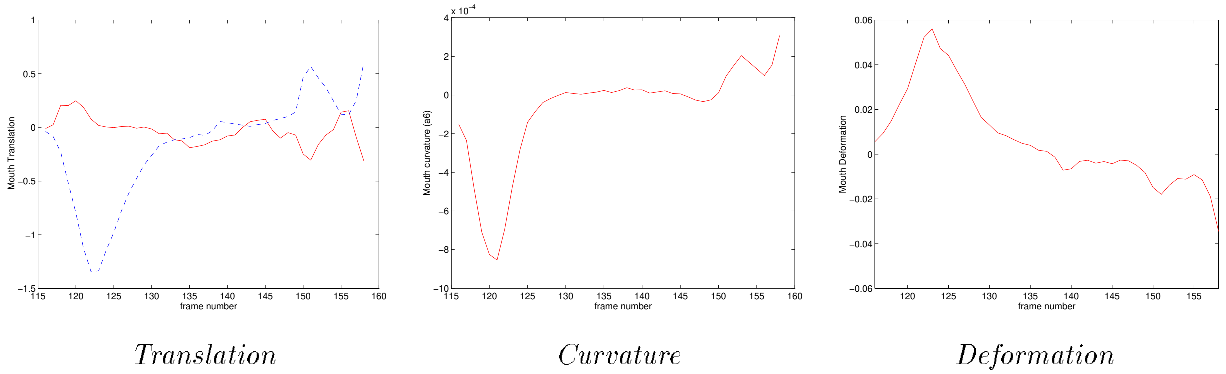


Figure 9: Smile mouth parameters. For translation, solid and dashed lines indicate horizontal and vertical motion respectively.

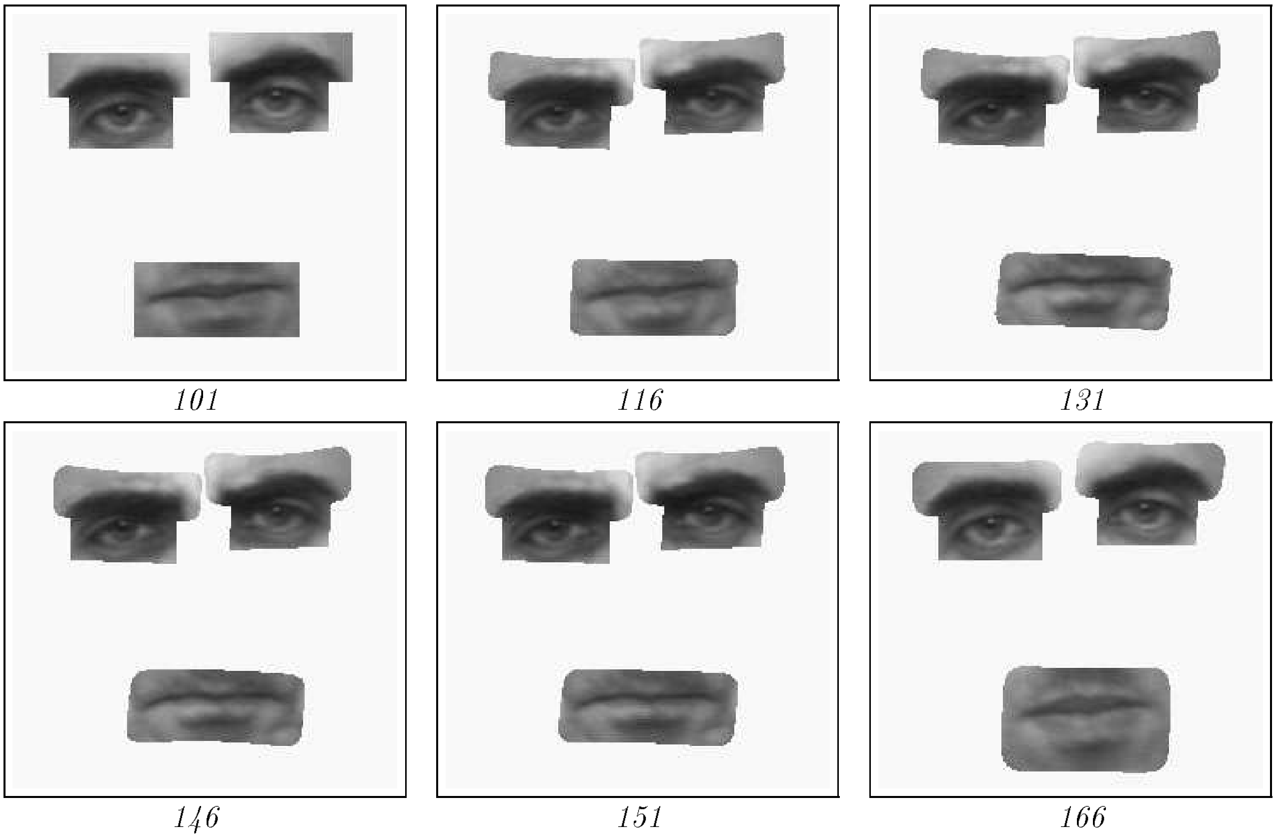


Figure 10: Anger experiment: facial expression tracking. Features every 15 frames.

c) and clearly shows the negative curvature corresponding to an upward bending motion at the initiation of the smile. The other significant cue to perceiving a 'smile' expression is a deformation of the mouth which resembles stretching in the horizontal direction; this is clearly visible in the plot of the mouth deformation ($a_1 - a_5$) on the right in the figure.

6.2 Anger experiment

An image sequence of an anger expression is shown in Figure 10. The anger expression (Figure 11) is characterized by a an initial pursing (or flattening) of the lips; then, in this case, a long slow downward curvature which ends abruptly around frame 150, after which the mouth curves and deforms back to the relaxed position. In addition to the mouth motion, the eyes and brows play a significant role. Figure 11 shows how the brows move together and down while becoming flatter (negative curvature) during the initiation of the expression. The nasal edges of the brows also dip downward causing opposite curl of the two brows.

These motions are reversed on cessation of the expression. In this case, only one of the eyes experiences the characteristic narrowing of anger, as indicated in the plot of deformation.

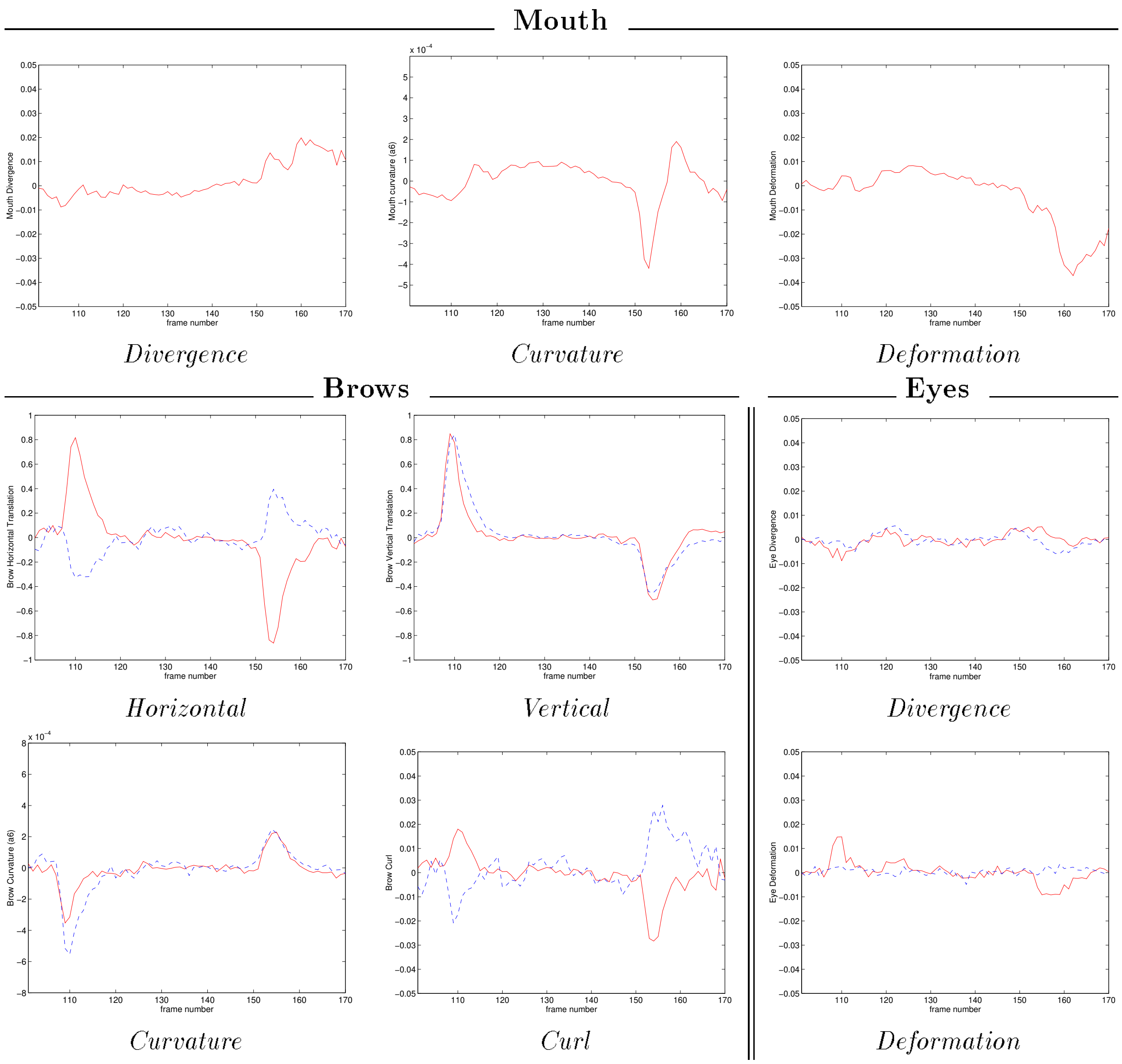


Figure 11: Anger motion parameters; the solid line indicates the right eye or brow while the dashed line indicates the left eye or brow.

6.3 Surprise experiment

Features tracked during a "surprise" expression are shown in Figure 12 and the significant parameters characterizing the expression are plotted in Figure 13. During the initiation of the expression the mouth translates down, diverges, and deforms significantly. Simultaneously, the brows and eyes move upward, the brows arch, and the eyes deform as they widen. The ending phase in this example shows a more gradual reversal of these parameters, returning the face to the resting position.

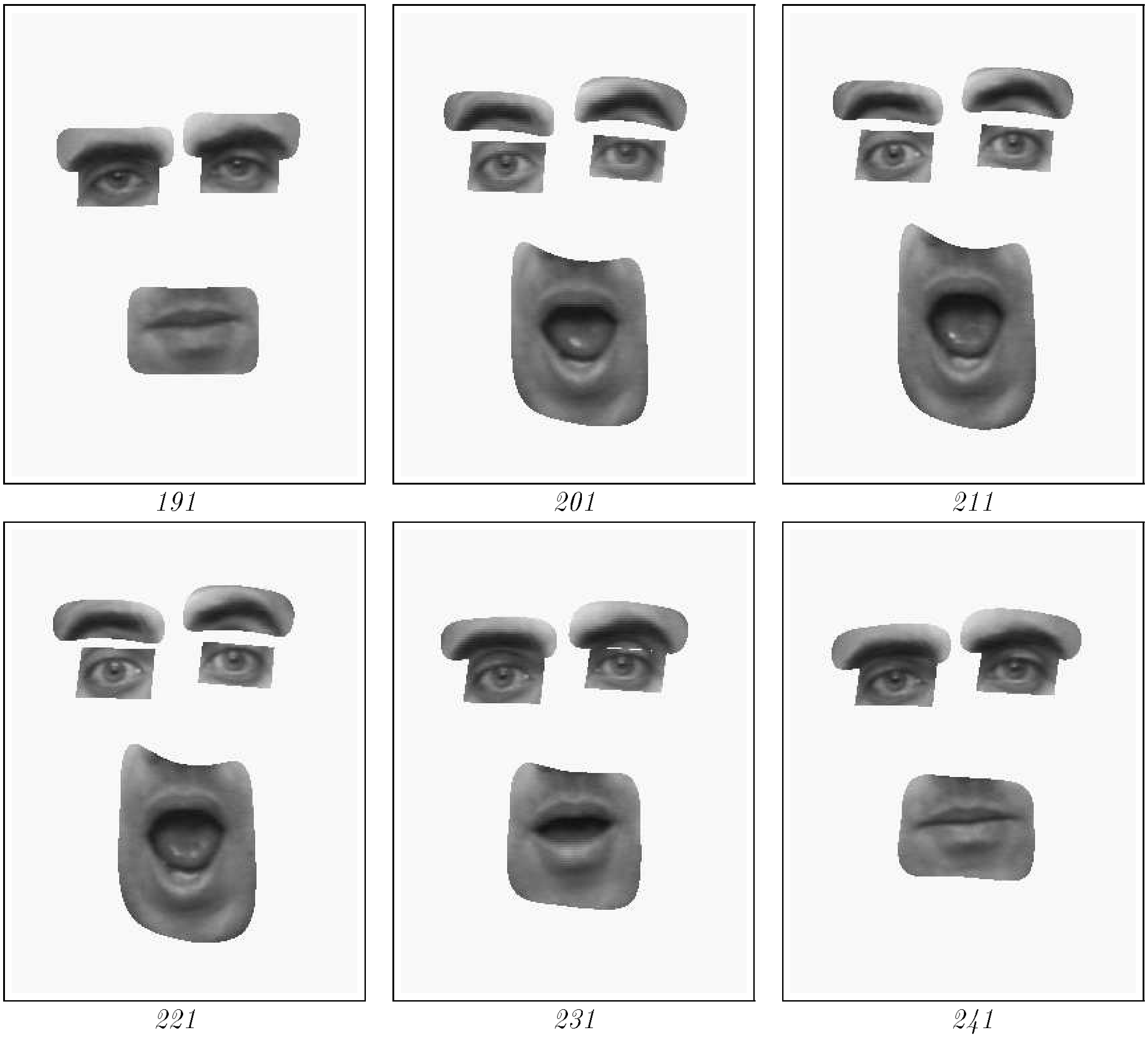


Figure 12: Surprise experiment: facial expression tracking. Features every 10 frames.

6.4 Blinking experiment

Figure 14 shows an image sequence in which the subject blinks twice. When the motion of the blinking eye is modeled as affine between two frames, the blinking is readily apparent

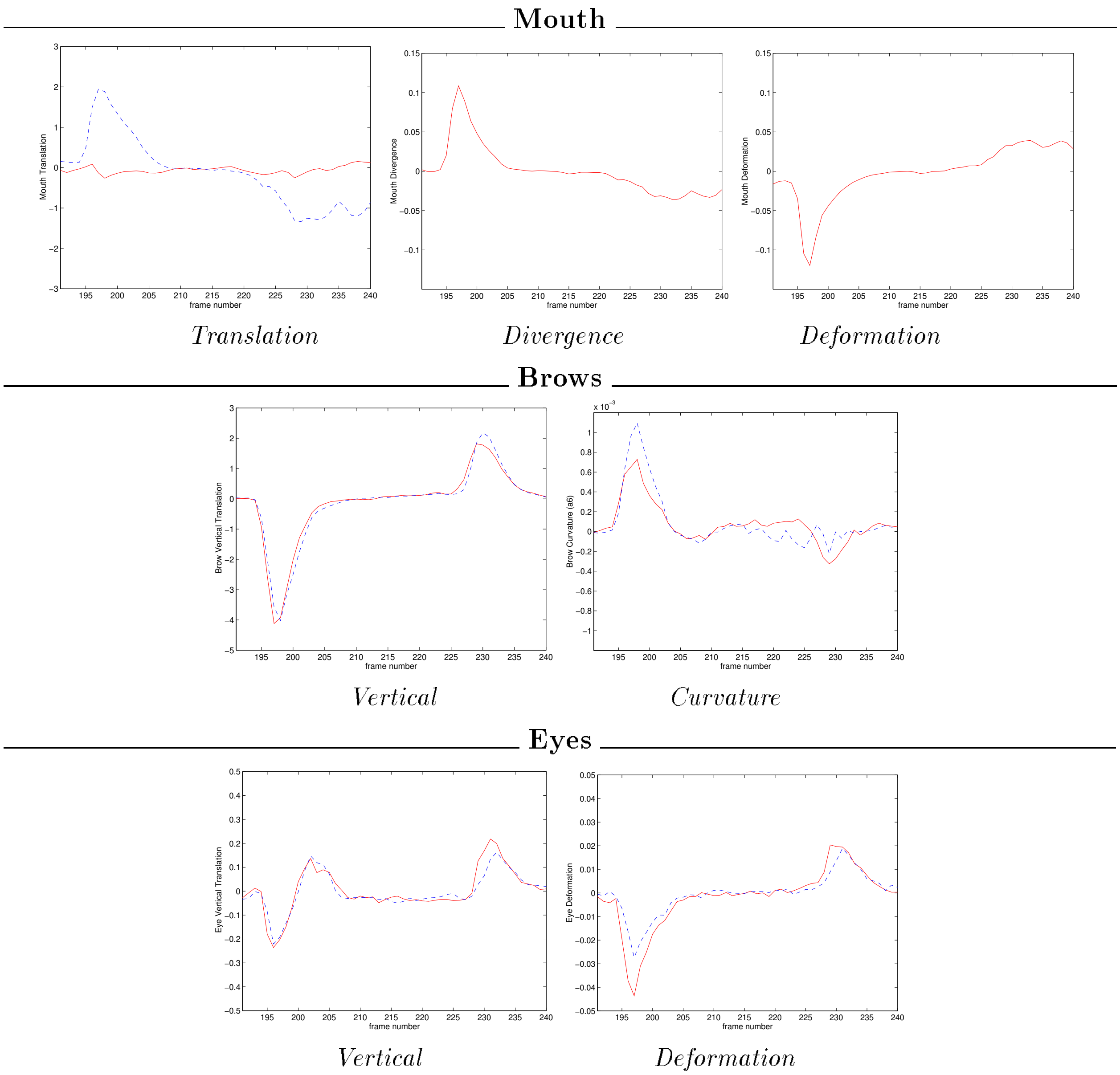


Figure 13: Surprise parameters. For the mouth translation, the solid line indicates horizontal motion while the dashed line indicates vertical motion. For the eye and brows, the solid and dashed lines indicate left and right respectively.

in the plots of vertical translation, divergence, and deformation. Figure 15 shows the rapid onset of a blink around frame 245. The blink subsides more gradually than it began and then another blink begins around frame 260 and likewise reverses more gradually. Notice that the tracked eye regions are not affected by the blinking action.

6.5 Looming experiment

The image sequence in Figure 16 illustrates facial expressions (smiling and surprise) in conjunction with rigid head motion (in this case looming). The figure plots the regions corre-

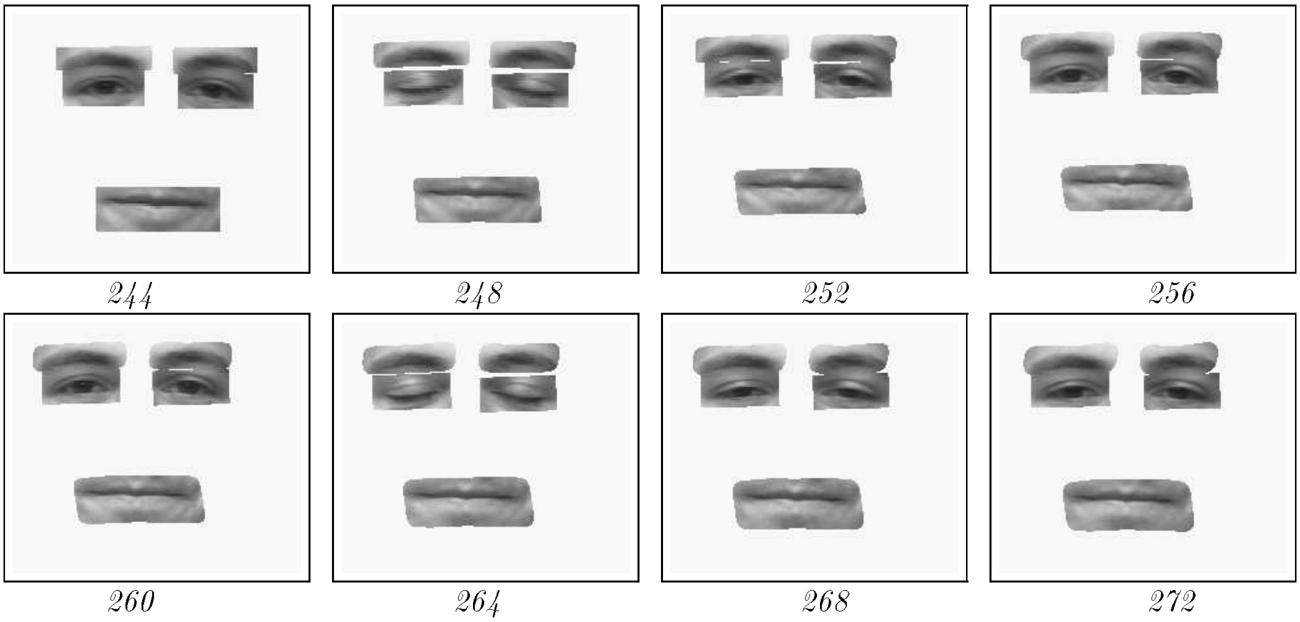
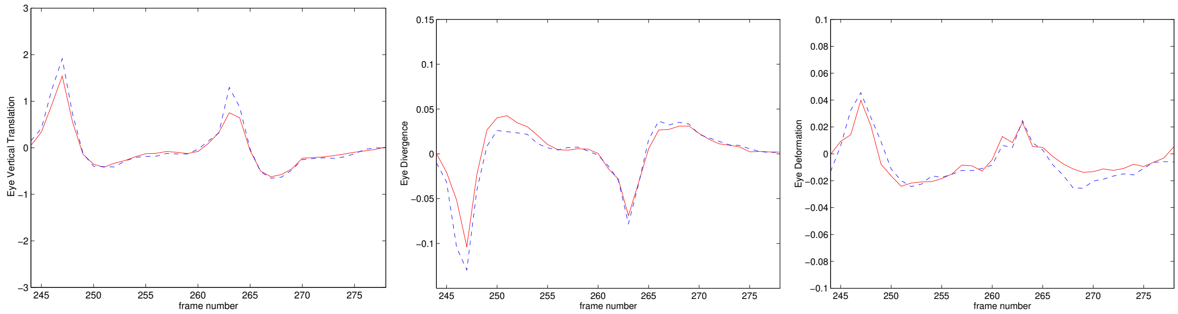


Figure 14: Blinking experiment: facial feature tracking. Features every four frames.



Translation *Divergence* *Deformation*

Figure 15: Blinking experiment: motion parameters for the eyes. Left eye = solid line, right eye = dashed line.

sponding to the face and the facial features tracked across the image sequence. The parameters describing the planar motion of the face are plotted in Figure 17 where the divergence due to the looming motion of the head is clearly visible in the plot of divergence. Analyzing the plots of the facial features in Figure 18 reveals that a smile expression begins around frame 125 with an increase in mouth curvature followed by a deformation of the mouth. The curvature decreases between frames 175 and 185 and then a surprise expression begins around frame 220 with vertical eyebrow motion, brow arching, and mouth deformation.

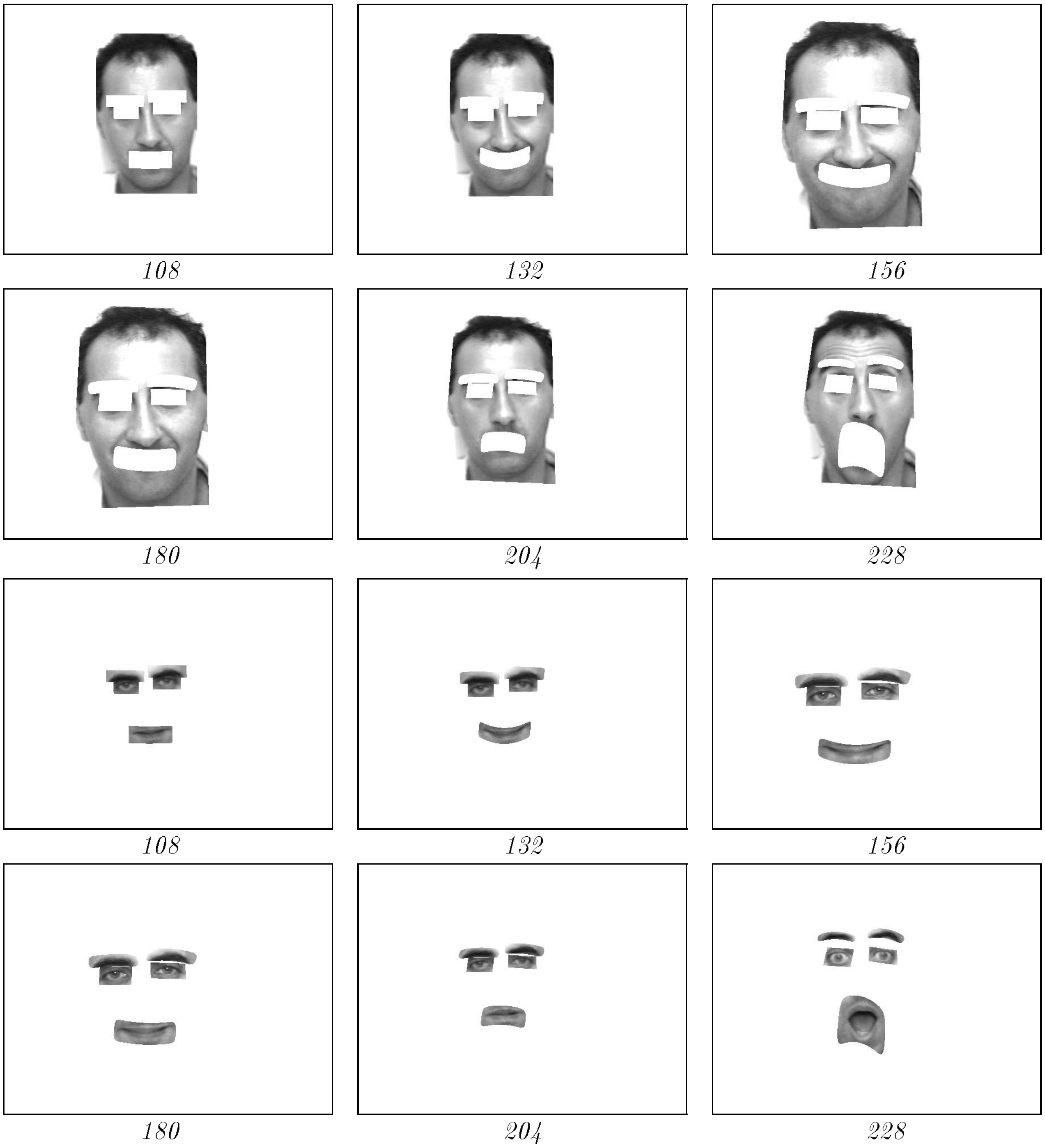


Figure 16: Looming experiment. Facial expression tracking with rigid head motion (every 24 frames).

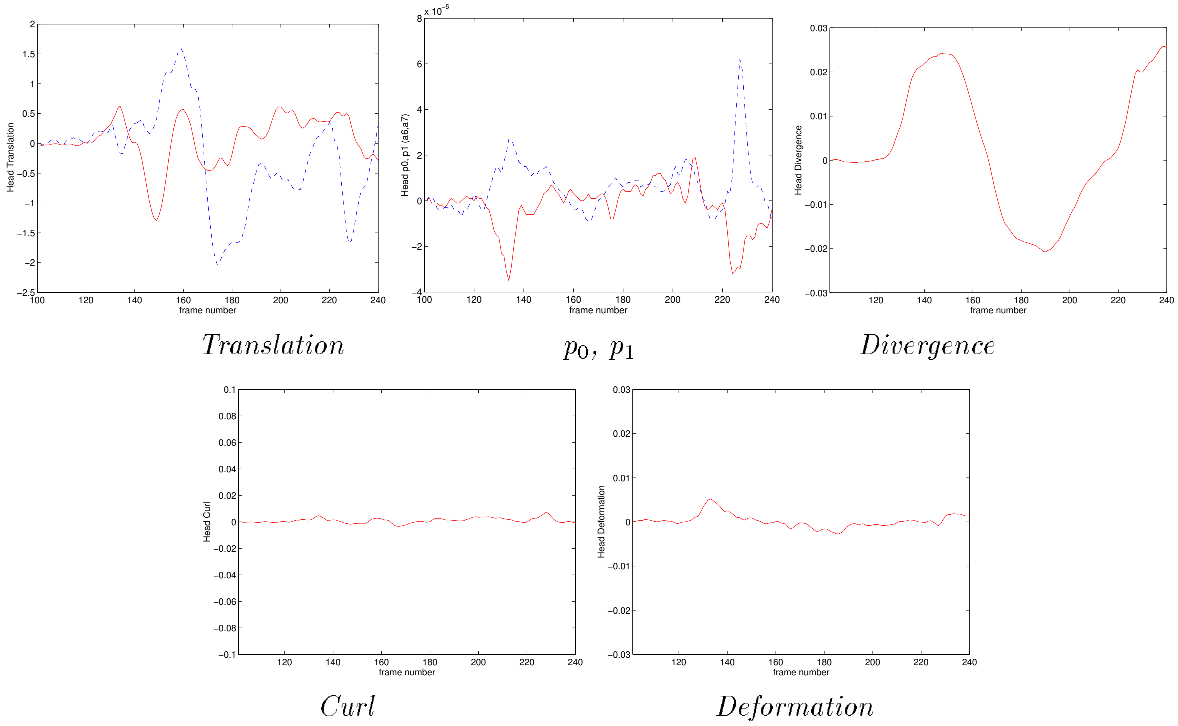


Figure 17: Looming face motion parameters. Translation: solid = horizontal, dashed = vertical. Quadratic terms: solid = p_0, dashed = p_1.

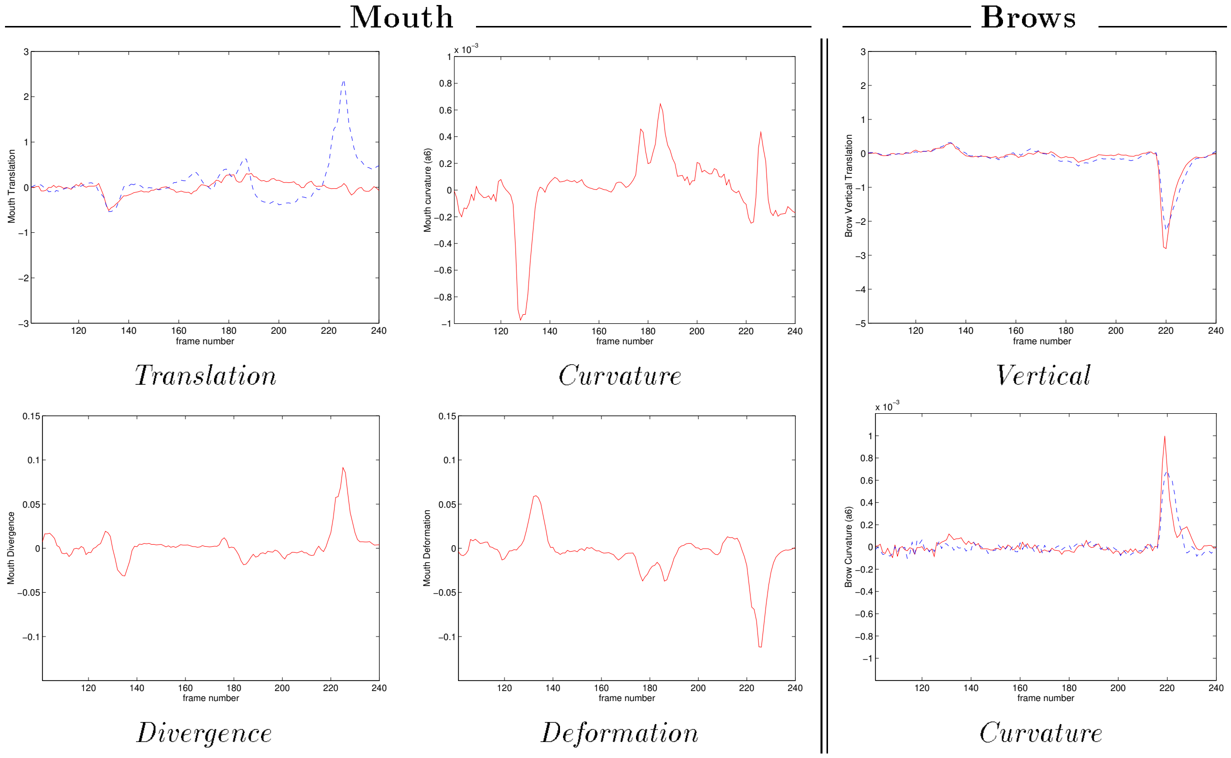


Figure 18: Looming motion parameters. Mouth translation: solid and dashed lines indicate horizontal and vertical motion respectively. For the brows, the solid and dashed lines indicate the left and right brows respectively.

6.6 Rotation experiment

Figure 19 illustrates a sequence with more complex rigid face motion due to rapid head rotations. In addition to the rapid motion, the sequence contains sections of motion blur, loss of focus, and saturation. Despite the low quality of the sequence the tracking of the head and features is robust.

In Figure 20 the plot of curl shows that the face rotates clockwise in the image plane, then rotates counterclockwise, pauses briefly, and continues the counterclockwise motion. The plot of p_0 (solid line) indicates that the head is rotating about the neck axis roughly in conjunction with the curl. To a lesser degree the head pitches fore and aft (p_1 – dashed line) and looms forward and back.

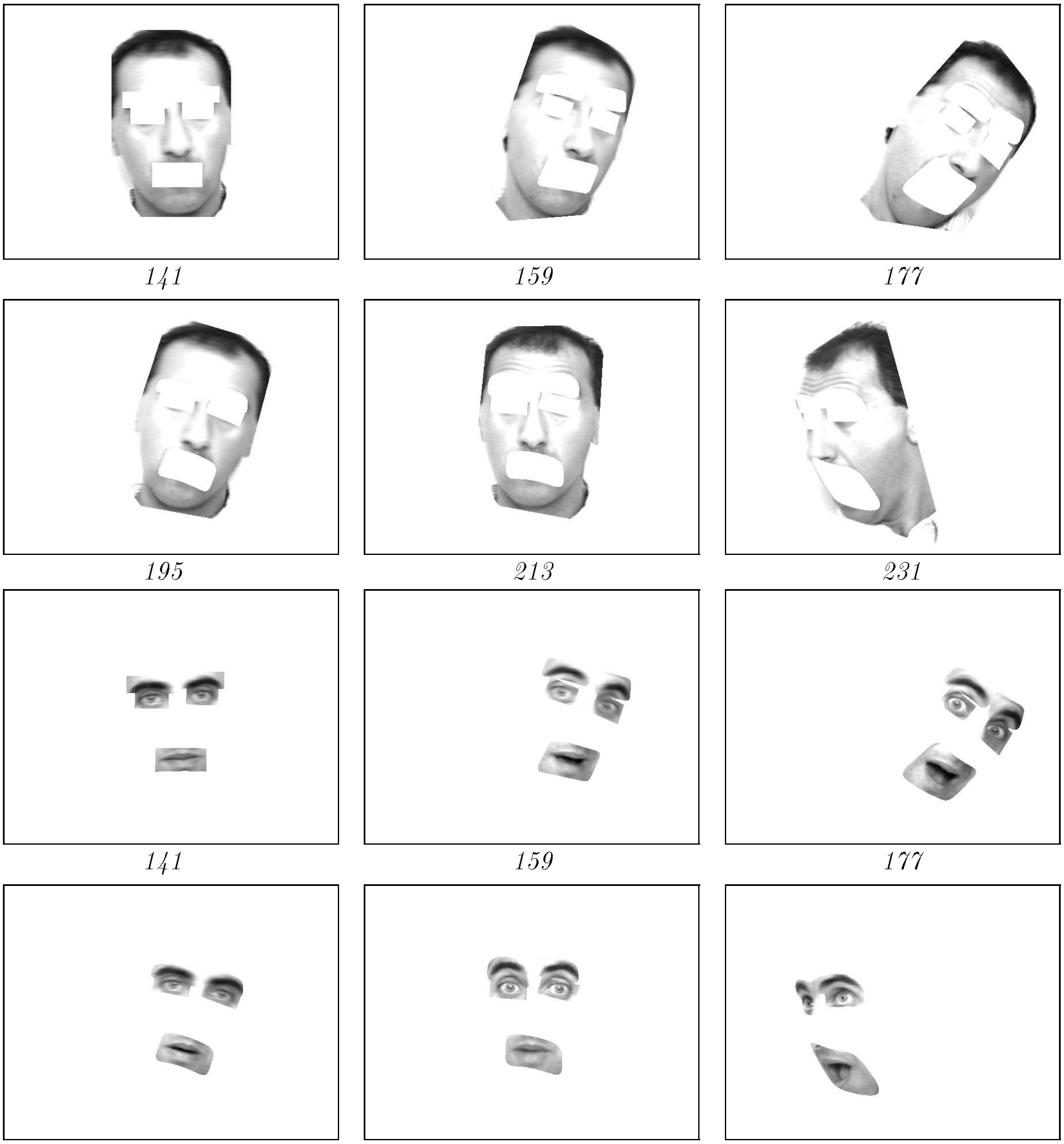
The facial features are consistently tracked despite the large image motions and the recovered parameters in Figure 21 indicate that two surprise expressions take place during the sequence as characterized by the eyebrow motion and mouth deformation (see the surprise example with a static head for comparison).

7 Mid- and High-Level Representations

The deformation and motion parameters computed in the previous section demonstrate that derivation of mid- and high-level descriptions of facial actions can be achievable robustly. In this section we discuss these representations.

The thresholds of the primary parameters described in Table 2 are set to filter out most of the small and noisy estimates of deformation and motion. The mid-level representation describes the observed facial changes in each frame. It is based on the predicates given in Table 3 for the 'mouth' (similar tables were developed for the eyebrows and eyes).

The mid-level representation that describes the head motions is given in Table 4. The planar model of facial motion is primarily used to stabilize the head motion so that the relative motions of the features may be estimated. The motion of this plane also provides a qualitative description of the head motion. For example, we can qualitatively recover whether the head is rotating or translating. To accurately recover the true 3D motion of the



141 159 177

195 213 231

141 159 177

195 213 231

Figure 19: Rotation experiment. Rigid head tracking, every 18th frame.

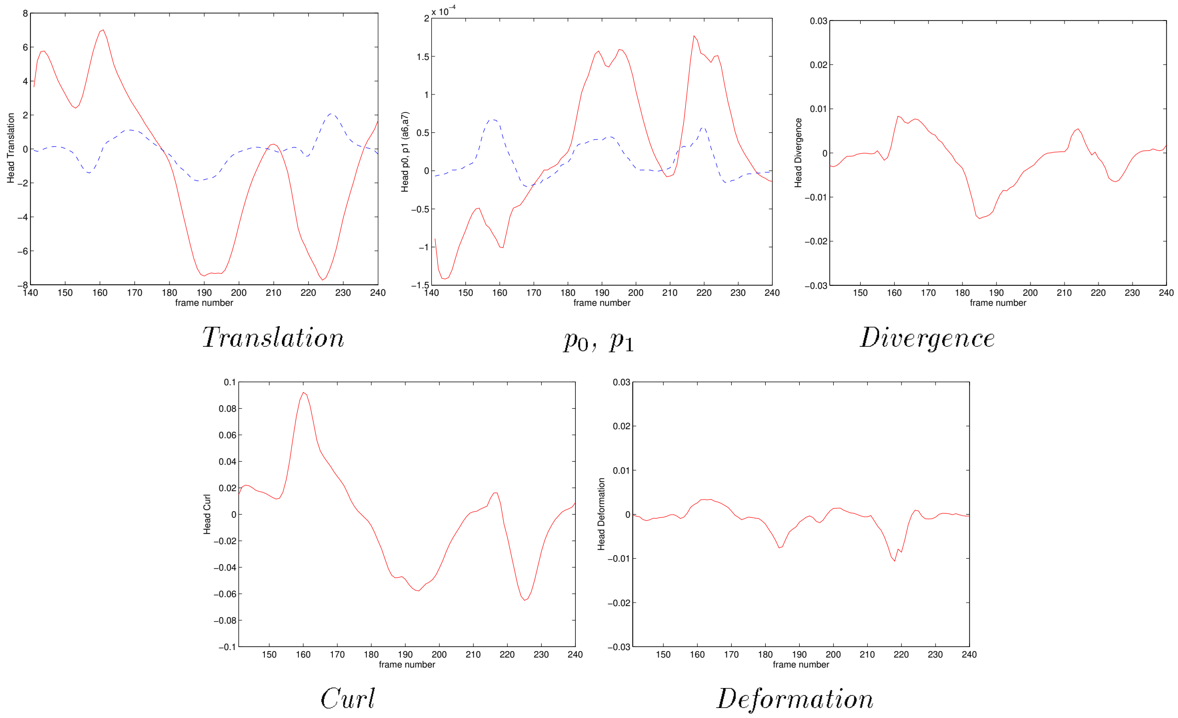


Figure 20: Rotate face motion parameters. Translation: solid = horizontal, dashed = vertical. Quadratic terms: solid = p_0, dashed = p_1.

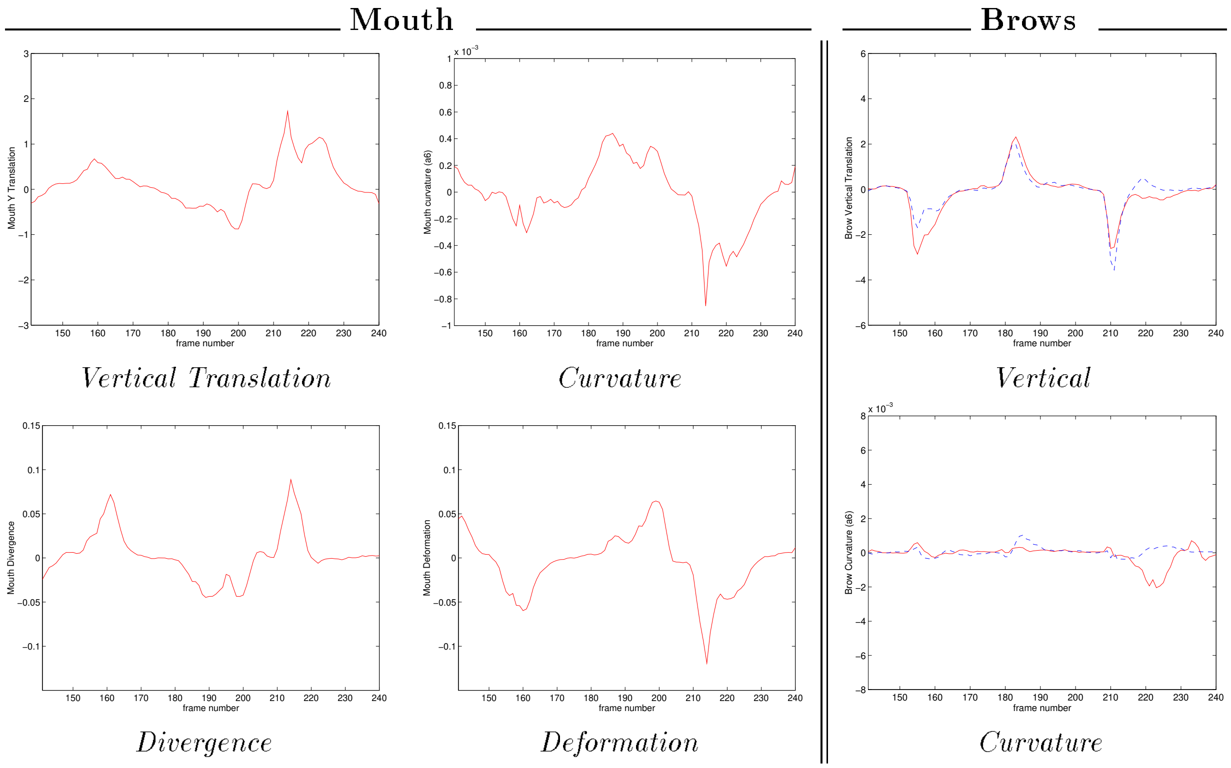


Figure 21: Rotate motion parameters. For the brows, the solid and dashed lines indicate left and right brows respectively.

Table 3: The mid-level predicates derived from deformation and motion parameter estimates.

Parameter	Threshold	Derived Predicates
a_0	> 0.25	Mouth rightward
	< -0.25	Mouth leftward
a_3	< -0.1	Mouth upward
	> 0.1	Mouth downward
Div	> 0.02	Mouth expansion
	< -0.02	Mouth contraction
Def	> 0.005	Mouth horizontal deformation
	< -0.005	Mouth vertical deformation
$Curl$	> 0.005	Mouth clockwise rotation
	< -0.005	Mouth counterclockwise rotation
c	< -0.0001	Mouth curving upward ('U' like)
	> 0.0001	Mouth curving downward

head would require a model more general than the planar assumption; this is part of our ongoing research.

Table 4: The mid-level predicates derived from deformation and motion parameter estimates as applied to head motion.

Parameter	Threshold	Derived Predicates
a_0	> 0.5	Head rightward
	< -0.5	Head leftward
a_3	< -0.5	Head upward
	> 0.5	Head downward
Div	> 0.01	Head expansion
	< -0.01	Head contraction
Def	> 0.01	Head horizontal deformation
	< -0.01	Head vertical deformation
$Curl$	> 0.005	Head clockwise rotation
	< -0.005	Head counterclockwise rotation
p_0	< -0.00005	Head rotating rightward around the neck
	> 0.00005	Head rotating leftward around the neck
p_1	< -0.00005	Head rotating forward
	> 0.00005	Head rotating backward

The high-level representation of facial actions (i.e., the facial expression recognition procedure) considers the temporal consistency of the mid-level predicates to minimize the effects of noise and inaccuracies in the motion and deformation models. The rules for detecting the beginnings and endings of expressions are given in Table 5. These rules are applied to the predicates of the mid-level representation. Generally, a beginning/ending has to be detectable continuously over at least four consecutive frames for the action to be recognized.

The high-level representation of head motion is currently limited to detecting backward and forward motions, right and left rotations around the neck, and looming. These recognized motions are illustrated in the experimental results found in the following section along with the facial expressions that are detected.

Table 5: The rules for classifying facial expressions (B=beginning, E=ending).

Expr.	B/E	Satisfactory actions
Anger	B	inward lowering of brows and mouth contraction
Anger	E	outward raising of brows and mouth expansion
Disgust	B	mouth horizontal expansion and lowering of brows
Disgust	E	mouth contraction and raising of brows
Happiness	B	upward curving of mouth and expansion or horizontal deformation
Happiness	E	downward curving of mouth and contraction or horizontal deformation
Surprise	B	raising brows and vertical expansion of mouth
Surprise	E	lowering brows and vertical contraction of mouth
Sadness	B	downward curving of mouth and upward-inward motion in inner parts of brows
Sadness	E	upward curving of mouth and downward-outward motion in inner parts of brows
Fear	B	expansion of mouth and raising-inwards inner parts of brows
Fear	E	contraction of mouth and lowering inner parts of brows

7.1 Resolving conflicts between expressions

The system is designed to identify and recognize facial expressions from long video clips (i.e., clips including 3-6 expressions). We simplified the behavioral model of our subjects by asking them to display one emotion at a time, and to include a short neutral state between expressions (we allowed, however, co-occurrences of 'smile' and 'surprise' since these were often displayed by our subjects). Since the six expression classifiers operate on the whole sequence independently, the system may create conflicting hypotheses.

Conflicts may arise when an ending of an expression is mistaken for the beginning of another expression. For example, the 'anger' recognition module may consider the lowering of the eyebrows during the ending of a 'surprise' expression as the beginning of an 'anger' expression. To resolve such conflicts, we employ a memory-based process that gives preference to the expression that started earlier.

8 Recognition Results

We carried out a large set of experiments to verify and evaluate the performance of the models proposed in this paper. The first set of experiments focused on the expressions of forty subjects who were asked to perform expressions in front of a digital camera. The second set of experiments involved digitizing video clips from television and movies. In the rest of this section we discuss the methodology used in acquiring the data and provide detailed statistics and analysis of the results.

8.1 Methodology

There are both technical difficulties in collecting data sets of sufficient spatial and temporal resolution and challenges in designing and interpreting experiments that evaluate the ability of a system to recognize expressions. Technically,

- image sequences should be sampled at 30Hz (or more) to minimize the magnitudes of rigid and non-rigid motions between consecutive images and provide for estimation of these motions,
- image resolution should allow the facial features to be of sufficient resolution to facilitate tracking and motion estimation, and
- the data that needs to be captured and processed is on the order of 10MByte/second, so that only short durations of expressions can be captured (on the order of 10 seconds of live video, assuming no compression is used).

Challenges associated with experimental design include

- expressions of emotions are hard to elicit in an artificial environment where people are not experiencing the typical associated emotions (see [11] for a discussion about how such experiments are conducted in psychology),
- determining what expression is "actually" being displayed is difficult because "different" expressions may appear quite similar, leading to variations in human recognition of expressions.

In an attempt to maintain some consistency in deciding which expression is actually being displayed, we used the the cues identified by psychological studies [4, 13] to determine the "ground truth" expressions.

We performed two sets of experiments. In the first, we recorded tens of expressions of forty subjects (having varied race, culture, and appearance) displaying their own choices of expressions. Our experimental subjects were asked to display emotional expressions without additional directions (in fact, we asked the subjects to choose any subset of the expressions and to display the expressions in any order and as naturally as they could). The expressions 'fear' and 'sadness' were hard to elicit compared to 'happiness,' 'surprise,' 'disgust' and 'anger.' The subjects were also asked to move their heads; but to avoid a profile view.

In the second set, we digitized tens of video clips recorded from talk shows, news programs, and movies. TV broadcasting, reception, video recording and digitization make the data quite noisy. As we will indicate later, interlacing in NTSC video creates additional problems.

While we asked our laboratory subjects not to speak while displaying expressions, the video recordings only occasionally include speechless expressions. The occurrence of speech affects the interpretation of the mouth motions since some motions due to speech may appear to be due to expressions. Therefore, we chose clips that are closely associated with emotional expressions (emotional behavior on talk shows, 'smiles' as substitutes for 'happiness,' etc.). The set of expressions recorded is dominated by 'smiles.'

8.2 Results of laboratory experiments

Figure 22 shows the 40 people that participated in our experiments. Our database of image sequences contains 70 sequences of each of these faces. Each sequence is about 9 seconds long and contains 1-3 expressions. Images are 560×420 pixels (taken at 30Hz).

On these 70 image sequences of 40 subjects displaying a total of 128 expressions, the system achieved a recognition rate of 95% for 'smile,' 91% for 'surprise,' 90% for 'anger,' 83% for 'fear,' 100% for 'sadness,' and 93% for 'disgust' (an overall average recognition rate of 92%).

Table 6 shows the details of our results. Occurrences of 'fear' and 'sadness' are less frequent than 'happiness,' 'surprise' and 'anger.' Some confusion of expressions occurred between the following pairs: 'fear' and 'surprise,' and 'anger' and 'disgust.' These distinctions rely on subtle coarse shape and motion clues that were not always accurately detected.

Table 6: Facial expression recognition results on forty subjects

Expression	Correct	False Alarm	Missed	Rate
Happiness	58	7	3	95%
Surprise	32	2	3	91%
Anger	18	-	2	90%
Disgust	14	2	1	93%
Fear	5	-	1	83%
Sadness	8	1	-	100%

The dynamic nature of facial expressions makes it difficult to demonstrate the experiments in print. Therefore, we provide selected images that will, hopefully, convey our results. Figure 23 shows four frames (taken as every fourth frame from the sequence) of the beginning of an 'anger' expression of a six year old boy. The text that appears on the left side of each image represents the mid-level predicates of the facial deformations, and the text that appears on the right side represents the mid-level predicates of the head motion. The text below each image displays the high-level description of the facial deformation and the head motion. Figure 24 shows the beginning of a 'smile' expression while the head is rotating initially leftward and then rightward.

Figure 22: The forty participants in the experiments.

Figure 23: Four frames (four frames apart) of the beginning of an 'anger' expression displayed by a six year old boy.

8.3 Results of video clips

We digitized 36 video clips of people expressing emotions (TV talk shows, news, movies, and commercials). Figure 25 shows representative frames from each of the 36 video clips. Variations in lighting, face size, and head orientation are visible.

Table 7 shows the details of our results on these 36 video clips. Only 'smiles' occurred with sufficient frequency to allow us to estimate classification accuracy.

Figure 26 shows four frames (every fourth frame from the sequence) of the beginning of a 'surprise' expression of a TV show host. Figure 27 shows four frames taken from the movie "Amadeus" in which Mozart displays 'fear.'

Figure 24: Four frames (four frames apart) of the beginning of a 'smile' expression.

8.4 Computational cost

With the exception of the rigid and non-rigid motion estimation computations, the algorithms described operate at near frame rates. The recovery of deformation and motion parameters requires about two minutes per frame on a Digital *Alpha* 3000 at full image resolution. We are currently studying possible speedups for the computations using an incremental estimation approach.

8.5 Observations from the experimental work

Our experiments uncovered several weaknesses in our models. Efforts are being directed towards improving the system to address these problems.

Figure 25: Representative frames from each of the 40 video clips

Table 7: Facial expression recognition results on 36 video clips

Expression	Correct	False Alarm	Missed	Rate
Happiness	35	4	2	95%
Surprise	6	1	1	86%
Anger	4	-	1	80%
Disgust	2	2	2	50%
Fear	1	-	-	100%
Sadness	3	1	2	60%

- The planar assumption used to model the face leads to large quantitative tracking inaccuracies for subjects with very spherical faces. Additionally, we want to analyze profile views of the head, but cannot currently track the motion of the head from the profile to frontal views using the planar model (assuming initial face viewing in profile).
- Various natural mouth motions (lip biting, and even simple opening) are sometimes mistakingly identified as 'smiles'. Also, speech remains a considerable source of false alarm expressions (usually, mouth openings are confused with 'smiles').
- Shape information is not directly integrated into our analysis, yet the use of shape can considerably improve the accuracy of the system at all levels. At the low level, tracking and feature deformation can benefit from feedback from a shape recognition module. At the high level the shape created by contours around the features can disambiguate expressions.

In addition, some technical observations:

- The size of the face in the image affects the accuracy and correctness of motion estimation. Considerable degradation occurs whenever faces occupy too small a portion of the image (about 1/8 of the NTSC standard). This limits somewhat the cases where facial expression analysis is applicable.
- Interlacing in broadcasting and very rapid body motions produce images which do not satisfy the basic assumption of brightness conservation inherent in motion estimation. The standard approach of processing only one of the interlaced fields at a time results in reduced image sizes and hence smaller features whose motion cannot be estimated as reliably.

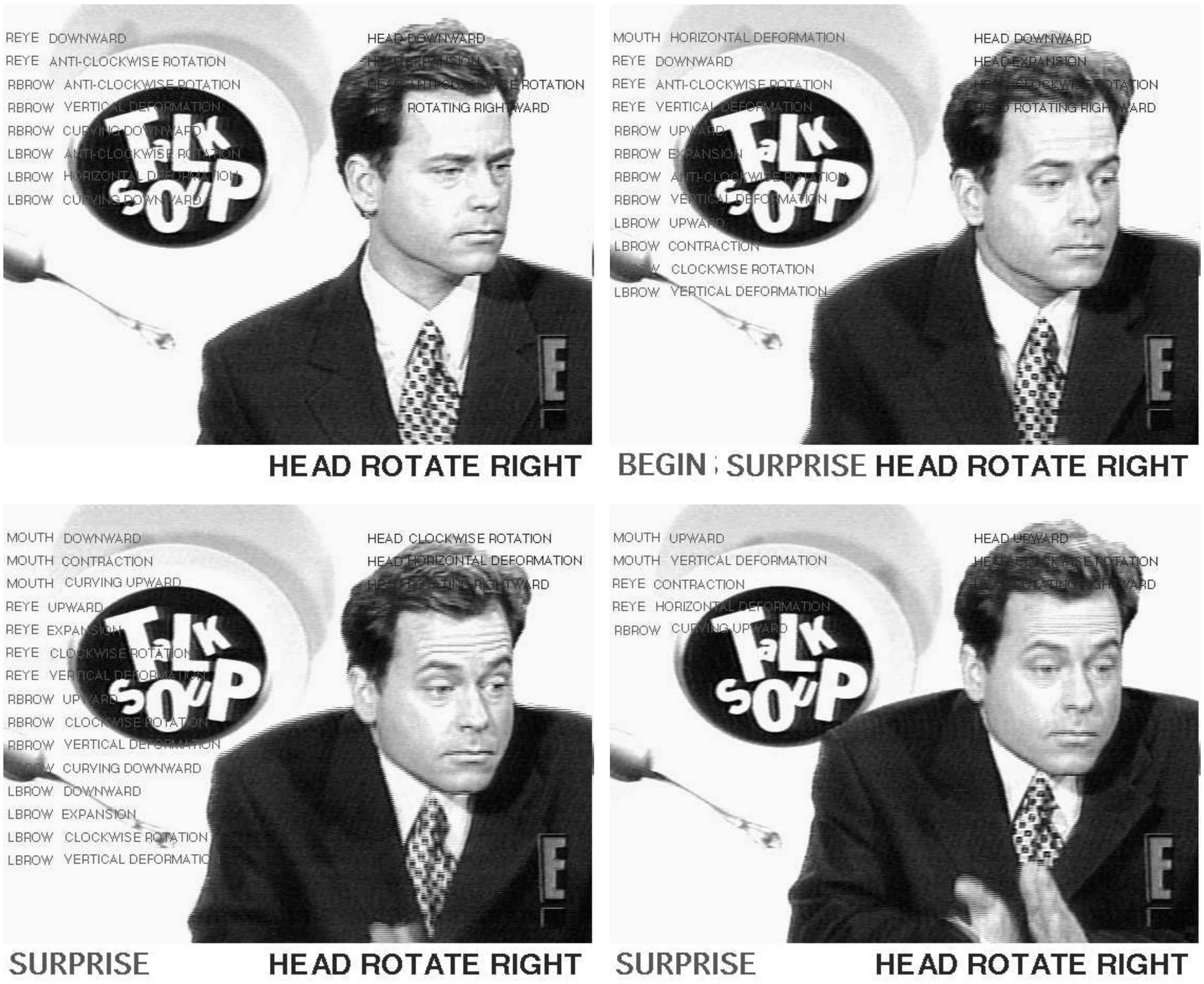


Figure 26: Four frames (four frames apart) of the beginning of a 'surprise' expression.

9 Conclusions and future research

In this paper we have proposed local parameterized models of image motion that can cope with the rigid and non-rigid facial motions that are an integral part of human behavior. Facial features are modeled locally to allow for accurate recovery of their deformations. In a series of experiments we have illustrated the effectiveness of these models in the presence of looming and considerable head rotation during facial deformation. Extensive experimentation with many subjects in natural situations, including television clips, indicates that expression recognition can be achieved accurately even in the presence of head motion.

Our ongoing work is focused on the modeling of coincident speech and facial expressions, more accurate recovery of 3D head motion, and dealing with a richer set of facial expressions.

Figure 27: Four frames (four frames apart) of the beginning of a 'fear' expression.

In particular we are exploring issues related to the integration of our motion-based approach with shape-based approaches. We are also pursuing the application of facial expression tracking and recognition to user interfaces.

Acknowledgements

We thank Rich Gold for the drawing of "Mr. Planar Head". We are thankful to Prof. Larry Davis for his valuable comments and suggestions. Without the cooperation of our forty volunteers in subjecting themselves to the intrusive experiments on their facial expressions this work would not have been possible. We are very grateful to them.

References

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(4):384–401, July 1985.

[2] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. In *Proc. Computer Vision and Pattern Recognition, CVPR-93*, pages 294–299, New York, NY, June 1993.

[3] A. Azarbayejani, T. Starner, B. Horowitz, and A. Pentland. Visually controled graphics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):602–604, June 1993.

[4] J.N. Bassili. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology*, 37:2049–2059, 1979.

[5] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In G. Sandini, editor, *Proc. of Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS Series*, pages 237–252. Springer-Verlag, May 1992.

[6] M. Black and P. Anandan. The robust estimation of multiple motions: Affine and piecewise-smooth flow fields. Technical Report P93-00104, Xerox PARC, December 1993.

[7] M.J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Proc. International Conference on Computer Vision, ICCV-93*, pages 231–236, Berlin, Germany, May 1993.

[8] M.J. Black and A. Jepson. Estimating multiple independent motions in segmented images using parametric models with local deformations. In *Proceedings of the Workshop on Motion of Non-rigid and Articulated Objects*, Austin, TX, November 1994.

[9] G. Chow and X. Li. Towards a system for automatic facial feature detection. *Pattern Recognition*, 26(12):1739–1755, 1993.

[10] R. Cipolla and A. Blake. Surface orientation and time to contact from image divergence and deformation. In G. Sandini, editor, *Proc. of Second European Conference on Computer Vision, ECCV-92*, volume 588 of *LNCS Series*, pages 187–202. Springer-Verlag, May 1992.

[11] P. Ekman, editor. *Emotion in the Human Face*. Cambridge University Press, 1982.

[12] P. Ekman. Facial expressions of emotion: An old controversy and new findings. *Philosophical Transactions of the Royal Society of London*, B(335):63–69, 1992.

[13] P. Ekman and W. Friesen. *Unmasking the Face*. Prentice-Hall, 1975.

[14] I. Essa, T. Darrell, and A. Pentland. Tracking facial motion. In *Proceedings of the Workshop on Motion of Non-rigid and Articulated Objects*, Austin, TX, November 1994.

[15] I.A. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. In *Proc. Computer Vision and Pattern Recognition, CVPR-94*, pages 76–83, Seattle, WA, June 1994.

[16] S. Geman and D.E. McClure. Statistical methods for tomographic image reconstruction. *Bulletin of the International Statistical Institute*, LII-4:5–21, 1987.

[17] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986.

[18] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proc. First International Conference on Computer Vision, ICCV-87*, pages 259–268, London, England, June 1987.

[19] J.J. Koenderink and A.J. van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22(9):773–791, 1975.

[20] H. Li, P. Roivainen, and R. Forcheimer. 3-D motion estimation in model-based facial image coding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):545–555, June 1993.

[21] M. Rosenblum, Y. Yacoob, and L.S. Davis. Human emotion recognition from motion using a radial basis function network architecture. In *Proceedings of the Workshop on Motion of Non-rigid and Articulated Objects*, Austin, TX, November 1994.

[22] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):569–579, June 1993.

[23] A.M. Waxman, B. Kamgar-Parsi, and M. Subbarao. Close-form solutions to image flow equations. In *Proc. First6 International Conference on Computer Vision, ICCV-87*, pages 12–24, London, England, June 1987.

[24] Y. Yacoob and L.S. Davis. Labeling of human face components from range data. In *Proc. Computer Vision and Pattern Recognition, CVPR-94*, pages 592–593, New York, NY, June 1993.

[25] Y. Yacoob and L.S. Davis. Computing spatio-temporal representations of human faces. In *Proc. Computer Vision and Pattern Recognition, CVPR-94*, pages 70–75, Seattle, WA, June 1994.

[26] A.W. Young and H.D.Ẽllis (Editors). *Handbook of Research on Face Processing.* Elsevier, 1989.

[27] A. Yuille and P. Hallinan. Deformable templates. In A. Blake and A. Yuille, editors, *Active Vision*, pages 21–38, MIT Press, Cambridge, MA, 1992.

[28] A.L. Yuille, D.S. Cohen, and P.W. Hallinan. Feature extraction from faces using deformable templates. In *Proc. Computer Vision and Pattern Recognition, CVPR-94*, pages 104–109, San Diego, CA, June 1989.