Tracking and Recognizing Rigid and Non-Rigid Facial Motions using Local Parametric Models of Image Motion

Michael J. Black
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
black@parc.xerox.com

Yaser Yacoob*
Computer Vision Laboratory
University of Maryland
College Park, MD 20742
yaser@cs.umd.edu

Abstract

This paper explores 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 motion of facial features during facial expressions and 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. The motion tracking and expression recognition approach performs with high accuracy in extensive laboratory experiments involving 40 subjects as well as in television and movie sequences.

1 Introduction

This paper describes a new method for recognizing human facial expressions in image sequences. There are three main contributions of this work. First, we describe a method for tracking rigid and non-rigid facial motions using a collection of local parameterized optical flow models. While parameterized models of image motion (for example affine) have become popular for the recovery of image motion in rigid scenes, their application to non-rigid and articulated motion is unconventional. Second, we show how the image-motion parameters can be interprated to recover high-level, semantic, descriptions of facial-feature motions. The interpretation of optical flow parameters in this way is novel and illustrates how the flow models concisely capture the relevant aspects of the feature motions. Finally, we demonstrate a system for recognizing facial expressions in image sequences containing significant head motion. The system has undergone extensive experimentation with over 100 video sequences of facial expressions gathered in both a lab setting and from television talk shows, news, movies, etc. These experiments indicate that our approach has a recognition accuracy of over 90%.

The approach can be thought of as having low, mid, and high levels. At the low level we take regions corresponding to the face, mouth, eyebrows, and eyes and model the rigid and non-rigid motions of these regions using a collection of parametrized flow models. For example, the face region, with the exception of the non-rigid features, is modeled as a plane. The motion of the plane is used to stabilize two frames of the image sequence and the motions of the features are then estimated relative to the stabilized face as illustrated in Figure 1. These feature motions are estimated

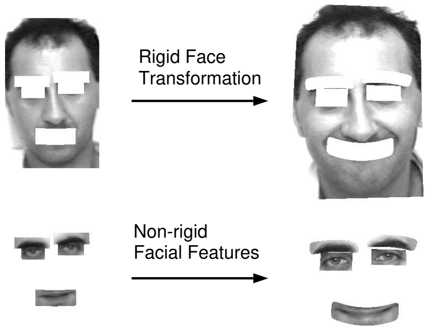

Figure 1: Illustration showing an example of a face undergoing a looming motion while smiling.

over an image sequence using a robust regression scheme [2] that makes the recovered motion parameters stable under adverse conditions such as motion blur, saturation, loss of focus, etc. In this paper we do not address the problem of initially locating the various facial features; this topic has been addressed in [9, 10].

The recovered image motion parameters correspond simply and intuitively to various facial expressions and are used to derive *mid-level predicates* describing the image motion of the facial features. *High-level recognition rules* describe the temporal structure of a facial expression in terms of these mid-level predicates. Currently we have implemented recognition rules for the six universal facial expres-

*The support of the Defense 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.

sions (surprise, sadness, anger, happiness, disgust and fear). We also recognize eye blinking and a variety of simple head motions.

Previous work. The recognition of facial expressions in image sequences with significant head motion is a challenging problem with many applications for human-computer interaction. Yet, while the coincidence of head and facial feature motion is prevalent in human behavior, it has so far attracted only little attention as a motion estimation problem. Previous work has typically focused on one part of the problem or the other: either rigid head tracking [1] with no facial expressions or expression recognition with either no motion at all [10] or a roughly stationary head with a changing expression [7, 8].

Models used in recognizing facial expressions vary in the amount of information about head shape and motion they contain. At one extreme are approaches which employ physically-based models of heads including skin and musculature [6, 7]. Slightly weaker models use deformable templates to represent feature shapes in the image [10]. Approaches that determine the expression by matching stored image templates to the current image [5] use even less explicit spatial information. At the other extreme is the work of Yacoob and Davis [8] in which facial expressions are recognized in image sequences 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. The piecewise parametric models of image motion explored here provide greater abstraction and robustness than the purely flow-based methods, are more general and robust than image template matching, and yet do not require detailed geometric or 3D information about face shape.

2 Models of Image Motion

Parameterized models of image motion make explicit the assumptions about the spatial variation of the optical flow within a region and typically assume that the flow can be represented by a low-order polynomial. Within small regions, an affine model of image motion is often sufficient

$$u(x, y) = a_0 + a_1 x + a_2 y \tag{1}$$
$$v(x, y) = a_3 + a_4 x + a_5 y \tag{2}$$

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 an 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:

$$\text{divergence} = a_1 + a_5 \tag{3}$$
$$\text{curl} = -a_2 + a_4 \tag{4}$$
$$\text{deformation} = a_1 - a_5. \tag{5}$$

Divergence, curl, and deformation, along with translation, will prove to be useful for describing facial expressions and are illustrated in Figure 2.

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:

$$u(x, y) = a_0 + a_1 x + a_2 y + p_0 x^2 + p_1 xy \tag{6}$$
$$v(x, y) = a_3 + a_4 x + a_5 y + p_0 xy + p_1 y^2 \tag{7}$$

where we have added two new terms p_0 and $p1$ to the affine model. These parameters roughly correspond to "yaw" and "pitch" respectively and are illustrated in Figure 2. While we have experimented with more complex models of rigid face motion we have found that this planar assumption is both simple and expressive enough to robustly represent rigid facial motions in a variety of situations.

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 [7] but their distributed nature does not admit simple, intuitive, characterizations of the motions as we saw above. Deformable templates on the other hand [10] encode information about shape but not 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. This can be achieved by adding new parameter, c, to the affine model

$$u(x, y) = a_0 + a_1 x + a_2 y \tag{8}$$
$$v(x, y) = a_3 + a_4 x + a_5 y + c x^2 \tag{9}$$

where c encodes curvature and is illustrated in Figure 2. This curvature parameter only captures the very coarse curvature of the features and cannot deal with asymmetric curvatures, however as the experiments will demonstrate, this seven parameter model captures the essential image motion of the mouth and eyebrows necessary for recognizing the six universal facial expressions.

Unfortunately this new curvature parameter is not invariant to head rotations. The curvature of the mouth and eyebrows should roughly be oriented with the principle axis of

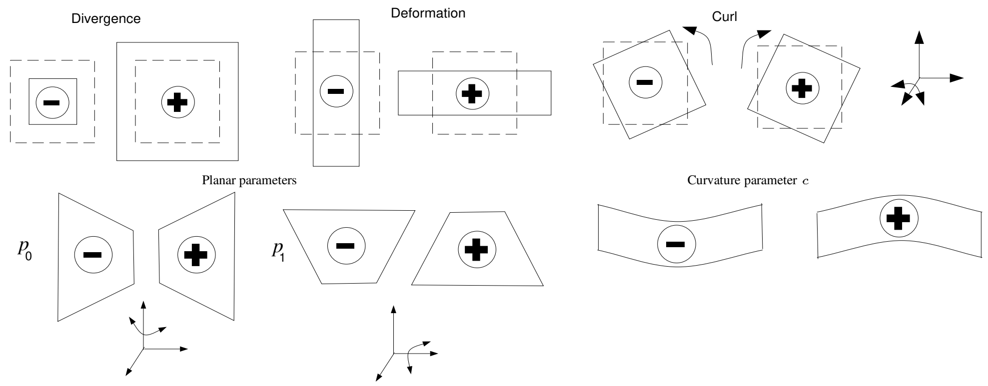

Figure 2: The figure illustrates the motion captured by the various parameters used to represent the motion of the regions. The solid lines indicate the deformed image region and the "–" and "+" indicate the sign of the quantity.

the face. To estimate the curvature with respect to the coordinate frame of the face we compute the orientation of the principle 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 the purpose of tracking.

2.1 Robust regression

To recover the parameters above, we employ a robust regression scheme [2]. For convenience of notation we define

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

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, respectively, the affine, planar, and affine+curvature flow models described above and $\mathbf{u} = [u, v]^T$.

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

$$I(\mathbf{x}, t) = I(\mathbf{x} - \mathbf{X}(\mathbf{x})\mathbf{P}_f, t+1), \quad \forall \mathbf{x} \in f, \tag{10}$$

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

$$\nabla I(\mathbf{X}(\mathbf{x})\mathbf{P}_f) + I_t = 0, \quad \forall \mathbf{x} \in f, \tag{11}$$

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

$$\sum_{\mathbf{x} \in f} \rho(\nabla I(\mathbf{X}(\mathbf{x})\mathbf{P}_f) + I_t, \sigma), \tag{12}$$

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

$$\rho(x, \sigma) = \frac{x^2}{\sigma + x^2}. \tag{13}$$

As the magnitudes of residuals $\nabla I(\mathbf{X}(\mathbf{x})\mathbf{P}_f) + I_t$ grow beyond a point their influence on the solution begins to decrease and the value of $\rho(\cdot)$ approaches a constant. The value σ is a scale parameter that effects the point at which the influence of outliers begins to decrease.

Equation 12 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 [2, 3] for details). The effect of this procedure is that initially no data are rejected as outliers then gradually the influence of outliers is reduced. To cope with large motions a coarse-to-fine strategy is used in which the motion is 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 in exactly the same way.

2.2 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. 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 eye 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 means that a simple quadrilateral 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 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

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

Exactly the same treatment is applied to the relative facial feature motions and these smoothed values are used for expression recognition. This simple tracking scheme works well for sequences of several hundred images with little accumulated error. The scheme also has the property of weighting current estimates more heavily than previous ones; this is appropriate for facial expressions which are typically of short duration. For longer image sequences, more sophisticated tracking schemes could be used; for example, Kalman filtering [1], segmentation information, and spatial constraints on the feature locations might be added.

3 Expression Recognition

The deformation and motion parameters described in the previous section can be used to derive mid- and high-level descriptions of facial actions; this section discusses these representations.

3.1 Mid-level representations

The parameters (such as translation and divergence) estimated for each feature are used to derive mid-level predicates that characterize the motion of the feature. The parameter values are first thresholded to filter out most of the small and noisy estimates. The mid-level representation describes the observed facial changes at each frame. Table 1 provides an example of the predicates 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 2. The planar model of facial motion is primarily used to stabilize the head motion so that the relative motion 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 when the head is rotating or translating. To accurately recover the true 3D motion of the head would require a model more general than the planar assumption.

Param.	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

Table 1: The mid-level predicates of the mouth derived from the motion parameter estimates.

Param.	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

Table 2: The mid-level predicates derived from the motion parameter estimates as applied to head motion.

3.2 High-level representations

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. In developing the high-level models we rely on the classification of facial expressions described in the psychological literature [4].

Following the temporal approach for recognition proposed in [8], we divide each facial expression into three temporal segments: the *beginning*, *apex* and *ending*. Figure 3 illustrates the temporal segments of a smile model. Notice that Figure 3 indicates that the change in parameter values

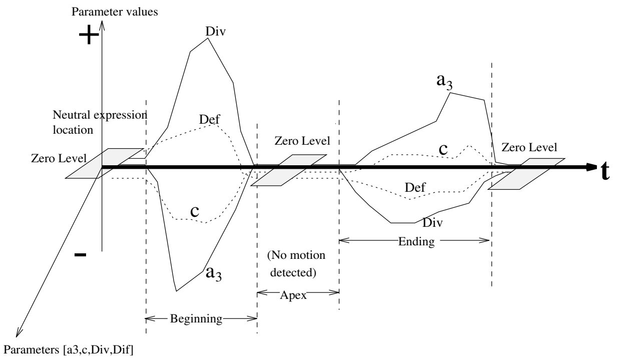

Figure 3: The temporal model of the 'smile.'

might not occur at the same frames at either the beginning or ending 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."

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.

The rules for detecting the beginning and ending of expressions are given in Table 3. These rules are applied to the predicates of the mid-level representation and are similar to those proposed in the psychological literature [4]. 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. The motion estimation and recognition behavior are illustrated in the remainder of the paper.

4 Motion Estimation Examples

The results of tracking the facial features during an 'anger' expression are shown in Figure 4 and some of the estimated motion parameters are shown in Figure 5. The 'anger' expression is characterized by an initial pursing (or flattening) of the lips then, in this case, a long slow downward curvature of the mouth that ends abruptly around frame 150 when the mouth curves and deforms back to the relaxed position. In addition to the mouth motion, the brows play a significant role. Figure 5 shows how the brows move together and down while becoming flatter (negative curvature) during the initiation of the expression. The nasal edges

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 & 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 & upward-inward motion in inner parts of brows
Sadness (E)	upward curving of mouth & downward-outward motion in inner parts of brows
Fear (B)	expansion of mouth & raising-inwards inner parts of brows
Fear (E)	contraction of mouth and lowering inner parts of brows

Table 3: The rules for classifying facial expressions (B = beginning, E = ending).

of the brows also dip downwards causing opposite curl (or rotation) for the two brows. These motions are reversed on cessation of the expression.

The image sequence in Figure 6 illustrates facial expressions ('smile' and 'surprise') in conjunction with rigid head motion (in this case looming). The figure plots the regions corresponding 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 7 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 8 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.

5 Recognition Results

We carried out a large set of experiments to verify and evaluate the performance of the recognition procedure 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.

Figure 9 shows the forty subjects who participated in our study; from these subjects we collected a database of 70 image sequences. Each sequence is about 9 seconds long and contains 1-3 expressions. Images are 560×420 pix-

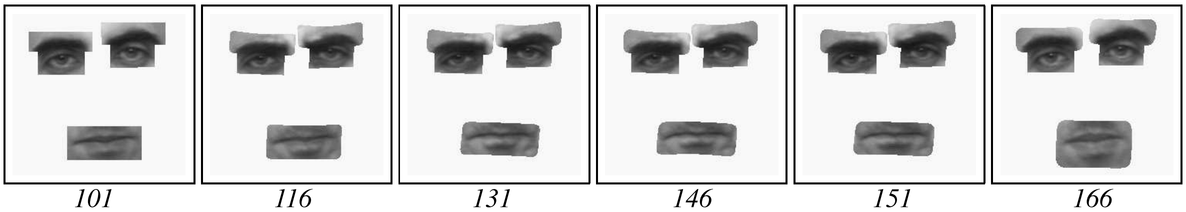

Figure 4: Anger Experiment: facial expression tracking. Features every 15 frames.

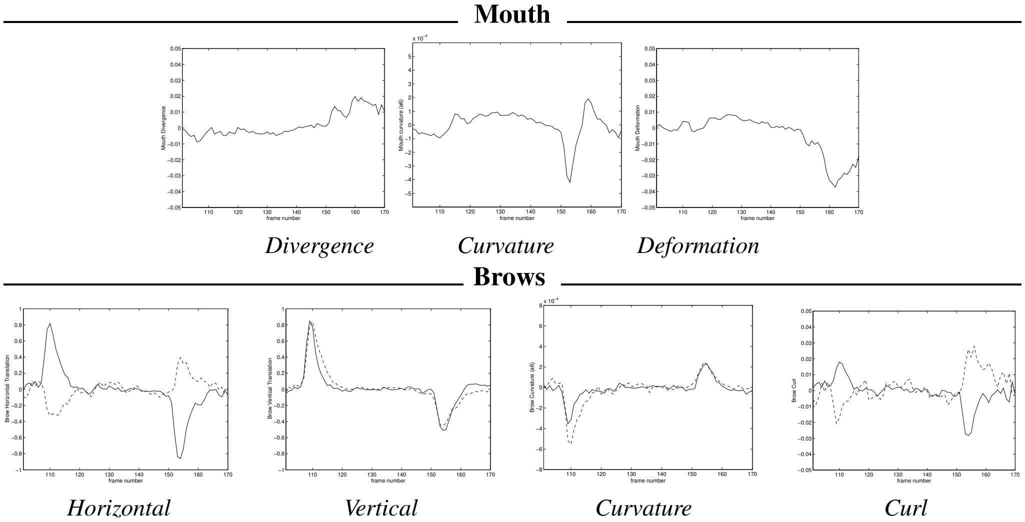

Figure 5: The Anger motion parameters; solid line indicates the right brow while the dashed line indicates the left brow.

els (taken at 30Hz). Ground truth for the data set was determined by inspection using the cues proposed in the psychological literature [4].

On these 70 image sequences 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 of 92%).

Table 4 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 information that were not always accurately detected.

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%

Table 4: Recognition results on forty subjects

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 10 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 deformations and the head motions. Figure 11 shows the beginning of a 'smile' expression while the head is rotating initially leftward and then rightward.

In a second set of experiments, we digitized 36 video-clips recorded from talk shows, news, and movies. TV broadcasting, reception, video-recording and digitization makes the data quite noisy. Table 5 shows the details of our results on these video-clips. Only 'smiles' occurred with sufficient frequency to allow us to estimate classification accuracy.

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%

Table 5: Recognition results on 36 video-clips

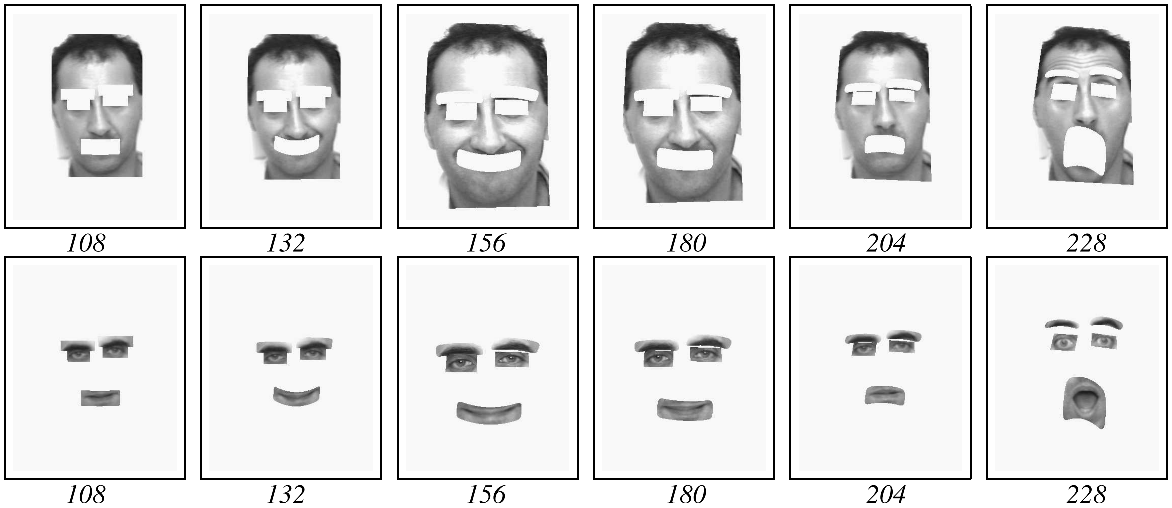

Figure 6: Looming Experiment. Facial expression tracking with rigid head motion (every 24 frames).

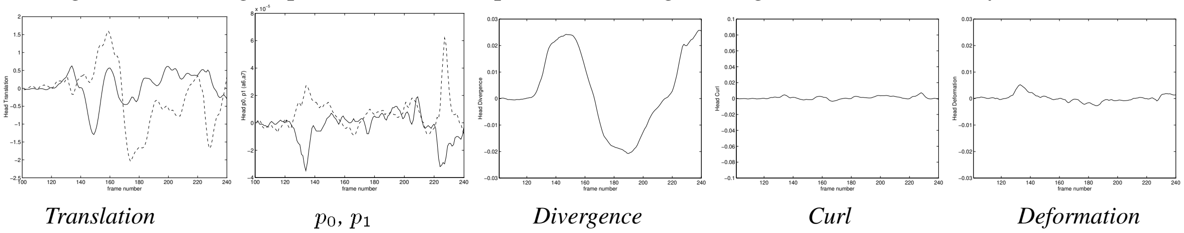

Figure 7: The Looming face motion parameters. Translation: solid = horizontal, dashed = vertical. Quadratic terms: solid = p_0, dashed = p_1.

Figure 10: Four frames (four frames apart) of the beginning of an 'anger' expression displayed by a six year old boy.

Figure 11: Four frames (four frames apart) of the beginning of a 'smile' expression.

6 Conclusion

In this paper we 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 rotations during facial deformations. 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.

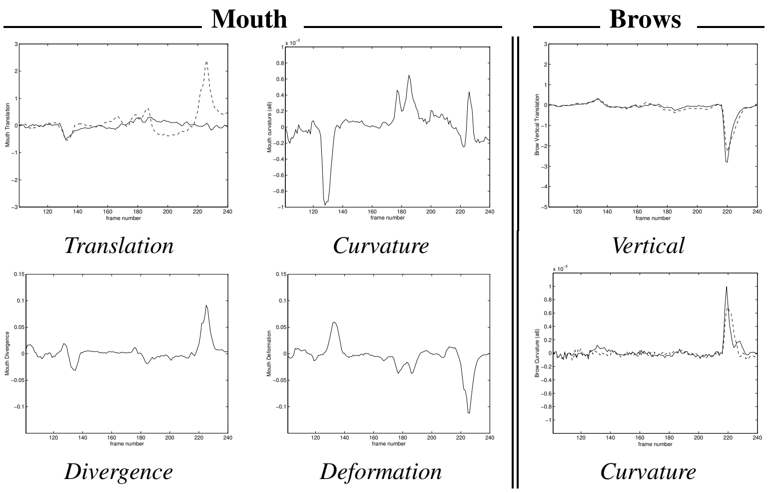

Figure 8: The Looming sequence. Mouth translation: solid and dashed lines indicate horizontal and vertical motion respectively. For the brows, the solid and dashed lines indicate left and right brows respectively.

Figure 9: The participants in the study.

Acknowledgements We thank Larry Davis for his valuable comments and suggestions. We also appreciate the cooperation of the forty volunteers who made our experiments possible.

References

[1] A. Azarbayejani, T. Starner, B. Horowitz, and A. Pentland. Visually controled graphics. *IEEE PAMI*, Vol. 15, No. 6, pp. 602–604, June 1993.

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

[3] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *ICCV-93*, pp. 231–236, Berlin, Germany, May 1993.

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

[5] I. Essa, T. Darrell, and A. Pentland. Tracking facial motion. In *Proc. Workshop on Motion of Non-rigid and Articulated Objects*, pp. 36–42, Austin, Texas, November 1994.

[6] I. A. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. In *CVPR-94*, pp. 76–83, Seattle, WA, June 1994.

[7] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE PAMI*, 15(6):569–579, June 1993.

[8] Y. Yacoob and L.S. Davis. Computing spatio-temporal representations of human faces. In *CVPR-94*, pp. 70–75, Seattle, WA, June 1994.

[9] Y. Yacoob and L.S. Davis. Labeling of human face components from range data. *CVGIP-Image Understanding*, 60(2):168–178, 1994.

[10] A.. L. Yuille, D. S. Cohen, and P. W. Hallinan. Feature extraction from faces using deformable templates. In *CVPR-89*, pp. 104–109, June 1989.