International Journal of Computer Vision, 8:2, 99–111 (1992)

# Feature Extraction from Faces Using Deformable Templates

ALAN L. YUILLE, PETER W. HALLINAN, AND DAVID S. COHEN
*Division of Applied Sciences, Harvard University, 29 Oxford St., Cambridge, MA 02138*

Received February 11, 1992.

**Abstract**

We propose a method for detecting and describing features of faces using deformable templates. The feature of interest, an eye for example, is described by a parameterized template. An energy function is defined which links edges, peaks, and valleys in the image intensity to corresponding properties of the template. The template then interacts dynamically with the image by altering its parameter values to minimize the energy function, thereby deforming itself to find the best fit. The final parameter values can be used as descriptors for the feature. We illustrate this method by showing deformable templates detecting eyes and mouths in real images. We demonstrate their ability for tracking features.

## 1 Introduction

The ability to detect and describe salient features is an important component of a face recognition system. Such features include the eyes, nose, mouth, and eyebrows. This task is hard despite pioneering work by Kanade (1977) and much research on edge detection and image segmentation. Current edge detectors, for example, seem unable to reliably find features such as the boundary of the eye. The problem seems to be that the edges of the eye rarely correspond to the idealized step edges in intensity assumed by most edge detectors. Moreover, even when local evidence for edges is available it is hard to organize this local information into a sensible global percept.

We propose a new method to detect such features by using deformable templates. These templates are specified by a set of parameters which enables a priori knowledge about the expected shape of the features to guide the detection process. The templates are flexible enough to be able to change their size, and other parameter values, so as to match themselves to the data. The final values of these parameters can be used to describe the features. The method should work despite variations in scale, tilt, and rotation of head, and lighting conditions. Variations of the parameters should allow the template to fit any normal instance of the feature.

The deformable templates interact with the image in a dynamic manner. An energy function is defined which contains terms attracting the template to salient features, such as peaks and valleys in the image intensity, edges (edges alone seem insufficient), and the intensity itself. The minimum of the energy function corresponds to the best fit with the image. The parameters of the template are then updated by steepest descent. This corresponds to following a path in parameter space, and contrasts with traditional methods of template matching which would involve sampling the parameter space to find the best match (and would be very expensive computationally). Changing these parameters corresponds to altering the position, orientation, size, and other properties of the template. The initial values of the parameters, which may be very different from the final values, are determined by preprocessing.

These deformable templates have some similarities to elastic deformable models (Burr 1981a, 1981b; Durbin and Willshaw 1987; Durbin, Szeliski, and Yuille 1988) and to snakes (Kass, Witkin, and Terzopolous 1987). These elastic models, however, do not contain the domain-specific structure we assume. Our work also is related to methods (Pentland 1987; Ayache et al. 1989) of representing geometric structures in terms of parameterized models and fitting them to depth data.

Our work was originally reported by Yuille, Cohen, and Hallinan (1989). More recently, deformable templates have been applied to facial features by Bennett and Craw (1991) and Shackleton and Welsh (1991). We have also applied similar techniques to medical images (Lipson et al. 1990). Since then we have become aware of other work on deformable templates, most notably the work of Grenander and his collaborators (Grenander, Chow, and Keenan 1991).

## 2 Preprocessing

The deformable templates act on three representations of the image, as well as on the image itself. These representations are chosen to extract properties of the image, such as peaks and valleys in the image intensity and places where the image intensity changes quickly. An additional representation could be added to describe textural properties. An advantage of using these representations is that the templates need only be specified in simple terms. For example we do not need to specify the intensity values on the iris, merely that the iris is a valley in the image intensity. Another advantage of using these representations is that they enable long-range interactions to occur.

These representations do not have to be very precise, and they can be calculated fairly simply. We have tried several methods to extract these features, including morphological filters (Maragos 1987; Serra 1982) and robust deformable templates for valleys and peaks (Hallinan 1991). The advantage of these methods for extracting the edge, valley, and peak fields is that they yield measures of the strengths of the features in question. This gives us three fields $\Psi_e(x, y)$, $\Psi_v(x, y)$ and $\Psi_p(x, y)$ representing the (positive) strengths of the edge, valley, and peak fields. For example, the edge field will be largest near edges in the image. We then smooth the fields by convolving them with an exponential function $\exp\{-\rho(x^2 + y^2)^{1/2}\}$. The smoothing enables the interactions to be effective over longer distances. This gives us three fields $\Phi_e(x, y)$, $\Phi_v(x, y)$ and $\Phi_p(x, y)$ where (using $*$ to denote convolution)

$$\begin{aligned}\Phi_e(x, y) &= e^{-\rho(x^2+y^2)^{1/2}} * \Psi_e(x, y)\\ \Phi_v(x, y) &= e^{-\rho(x^2+y^2)^{1/2}} * \Psi_v(x, y)\\ \Phi_p(x, y) &= e^{-\rho(x^2+y^2)^{1/2}} * \Psi_p(x, y) \qquad (1)\end{aligned}$$

There is an additional field $\Phi_i(x, y)$ corresponding to the image intensity $I(x, y)$ itself. Examples of these fields extracted using morphological operations can be seen in figures 2, 7, and 10.

Introducing these potential fields will enable (by the interactions specified in section 3.1) strong edges, valleys, or peaks to attract objects a large distance away. This is an advantage of working on representations of the image rather than on the image itself. A final refinement stage acting directly on the image may perform the final alignment.

## 3 The Eye Template

After some experimentation and informal psychophysics on the salience of different features of eyes we decided that the template should consist of the following features:

1. A circle of radius $r$, centered on a point $\vec{x}_c$. This corresponds to the boundary between the iris and the whites of the eye and is attracted to edges in the image intensity. The interior of the circle is attracted to valleys, or low values, in the image intensity.
2. A bounding contour of the eye attracted to edges. This contour is modeled by two parabolic sections representing the upper and lower parts of the boundary. It has a center $\vec{x}_e$, width $2b$, maximum height $a$ of the boundary above the center, maximum height $c$ of the boundary below the center, and an angle of orientation $\theta$.
3. Two points, corresponding to the centers of the whites of the eyes, which are attracted to peaks in the image intensity. These points are labeled by $\vec{x}_e + p_1(\cos\theta, \sin\theta)$ and $\vec{x}_e + p_2(\cos\theta, \sin\theta)$, where $p_1 \geq 0$ and $p_2 \leq 0$. The point $\vec{x}_e$ lies at the center of the eye and $\theta$ corresponds to the orientation of the eye.
4. The regions between the bounding contour and the iris also correspond to the whites of the eyes. They will be attracted to large values in the image intensity. These components are linked together by three types of forces: (i) forces which encourage $\vec{x}_c$ and $\vec{x}_e$ to be close together, (ii) forces which make the width $2b$ of the eye roughly four times the radius $r$ of the iris, and (iii) forces which encourage the centers of the whites of the eyes to be roughly midway from the center of the eye to the boundary.

The template is illustrated in figure 1. It has a total of eleven parameters represented by $\vec{g} = (\vec{x}_c, \vec{x}_e, p_1, p_2, r, a, b, c, \theta)$. All of these are allowed to vary during the matching.

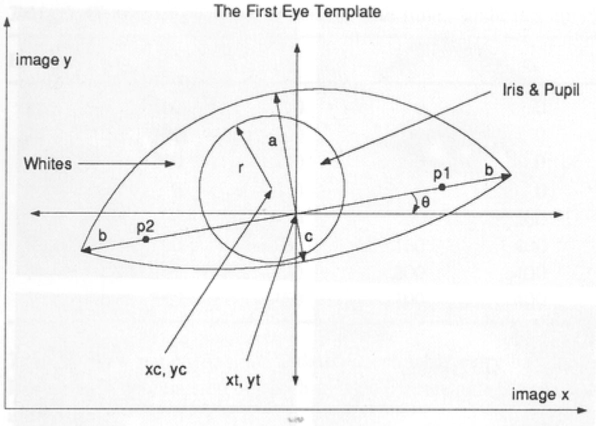

*Fig. 1.* A deformable template for an "archetypal" human eye, figure 2. It is parameterized by $a$, $b$, $c$, $x_t$, $y_t$, $x_c$, $y_c$, $r$, $\theta$, $p_1$, and $p_2$. $R_w$ and $R_b$ are intensity regions containing the whites and dark center of the eye respectively. $R_w$ is bounded by parabolic curves $\partial R_w$ specified by parameters $a$, $b$, and $c$. $R_b$ is bounded by a circle $\partial R_b$ of radius $r$.

To give the explicit representation for the boundary we first define two unit vectors

$$\vec{e}_1 = (\cos\theta, \sin\theta) \text{ and } \vec{e}_2 = (-\sin\theta, \cos\theta) \quad (2)$$

which change as the orientation of the eye changes. A point $\vec{x}$ in space can be represented by $(x_1, x_2)$ where $\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2$. Using these coordinates the top half of the boundary can be represented by a section of a parabola with $x_1 \in [-b, b]$

$$x_2 = a - \frac{a}{b^2}x_1^2 \quad (3)$$

Note that the maximal height, $x_2$, of the parabola is $a$ and the height is zero at $x_1 = \pm b$. Similarly the lower half of the boundary is given by

$$x_2 = -c + \frac{c}{b^2}x_1^2 \quad (4)$$

where $x_1 \in [-b, b]$.

### 3.1 The Energy Function for the Eye Template

We now define a potential energy function for the image which will be minimized as a function of the parameters of the template. This energy function not only ensures that the algorithm will converge, by acting as a Lyapunov function, but also gives a measure of the goodness of fit of the template.

The complete energy function $E_c(\vec{g})$ is given as a combination of terms due to valley, edge, peak, image, and internal potentials. More precisely,

$$E_c = E_v + E_e + E_i + E_p + E_{prior} \quad (5)$$

where: (i) The valley potentials are given by the integral over the interior of the circle divided by the area of the circle,

$$E_v = -\frac{c_1}{|R_c|}\int_{R_c} \Phi_v(\vec{x})dA \quad (6)$$

(ii) The edge potentials are given by the integrals over the boundaries of the circle divided by its length and over the parabolae divided by their lengths,

$$E_e = -\frac{c_2}{|\partial R_w|}\int_{\partial R_w} \Phi_e(\vec{x})\,ds - \frac{c_3}{|\partial R_w|}\int_{\partial R_w} \Phi_e(\vec{x})\,ds \quad (7)$$

(iii) The image potentials have contributions that attempt to minimize the total brightness inside the circle divided by its area,

$$E_i = \frac{c_4}{|R_w|}\int_{R_w} \Phi_i(\vec{x})\,dA \quad (8)$$

and maximize it between the circle and the parabolae (again divided by the area),

$$E_i = \frac{-c_5}{|R_w|}\int_{R_w} \Phi_i(\vec{x})\,dA \quad (9)$$

(iv) The peak potentials, evaluated at the two peak points, are given by

$$E_p = c_6\{\Phi(\vec{x}_e + p_1\vec{e}_1) + \Phi(\vec{x}_e + p_2\vec{e}_1)\} \quad (10)$$

(v) The prior potentials are given by

$$E_{\text{prior}} = \frac{k_1}{2}\|\vec{x}_e - \vec{x}_c\|^2 + \frac{k_2}{2}(p_1 - p_2 - \{r + b\})^2 + \frac{k_3}{2}(b - 2r)^2 + \frac{k_4}{2}(b - 2a)^2 + (a - 2c)^2) \quad (11)$$

Here $R_b$, $R_w$, $\partial R_b$, and $\partial R_w$ correspond to the iris, the whites of the eye, and their boundaries (see figure 1). Their areas, or lengths, are given by $|R_w|$, $|R_b|$, $|\partial R_w|$, and $|\partial R_b|$. $A$ and $s$ correspond to area and arc-length respectively.

## 4 The Algorithm and Simulation Results for Eyes

The algorithm uses a search strategy, based on steepest descent, that attempts to find the most salient parts of the eye in order. It first uses the valley potential to find the iris, then the peaks to orient the template, and so on.

To implement this strategy we divide the search into a number of epochs with different values of the parameters $\{c_i\}$ and $\{k_i\}$. The updating in each epoch is done by steepest descent in the total energy $E = E_v + E_p + E_e + E_i + E_{\text{prior}}$. The energy terms are written as explicit functions of the parameter values. For example, the sum over the boundary can be expressed as an integral function of $\vec{x}_e$, $a$, $b$, $c$, and $\theta$ by

$$\frac{1}{|\partial R_w|}\int_{\partial R_w} \Phi_e(\vec{x})\, ds$$

$$= \frac{c_3}{L(a,b)}\int_{x_1=-b}^{x_2=b} \Phi_e(\vec{x}_e + x_1\vec{e}_1 + \{a - \frac{a}{b^2}x_1^2\}\vec{e}_2)\, ds$$

$$+ \frac{c_3}{L(a,\, b)}\int_{x_1=-b}^{x_2=b} \Phi_e(\vec{x}_e + x_1\vec{e}_1 - \{c - \frac{c}{b^2}x_1^2\}\vec{e}_2)\, ds \tag{12}$$

where $s$ corresponds to the arc length of the curves and $L(a,\, b)$ and $L(c,\, b)$ to their total length.

The parameter values are updated by steepest descent, that is,

$$\frac{dr}{dt} = -\frac{\partial E}{\partial r}$$

It is assumed that preprocessing, or interactions between different templates, will allow the eye-template to start relatively near the correct position. See section 7 for a discussion of this.

This theory was tested on real images using a SUN4 computer. The valleys, peaks, and edges are first extracted and smoothed. On a typical eye, this gives rise to the potential fields shown in figure 2. The eye template is then given initial parameter values, positioned in the image and allowed to deform itself using the update equations.

Some initial experimentation was needed to find good values for the coefficients and a number of problems arose. For example, the intensity and valley terms over the circle attempt to find the maximum value of the potential terms *averaged* inside the circle. This led to the circle shrinking to a point at the darkest part of the iris. This effect could be countered by strengthening the edge terms, which pull the circle out to the edge between the iris and the whites of the eye. Another problem arose because the iris might also be partially hidden by the boundary of the eye, thus the part of the circle outside the boundary cannot be allowed to interact with the image. This can be dealt with by only considering the area of the circle inside the bounding parabolas.

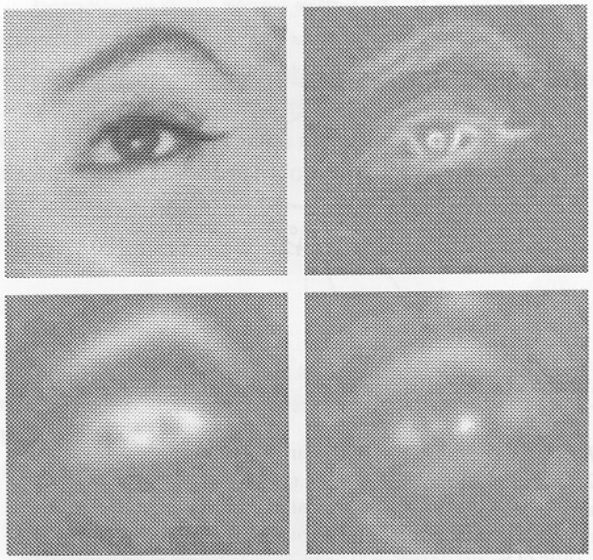

*Fig. 2.* The potential fields for an "archetypal" eye. The eye, top left, has edge, valley, and peak fields shown top right, bottom left, and bottom right. The strengths of the fields are shown in greytone, white is strong and black is weak.

The system worked well after good values were found for the coefficients. The templates usually converged to the eye provided they were started at or below it. The valleys from the eyebrows caused problems if the template was started above the eye.

The values of the coefficients changed automatically during the course of the program to define eight distinct epochs: Details of the implementation can be found in tables 1, 2, and 3, but note that the parameter values given there are not the only ones possible.

1. Only the valley forces are allowed to act on the template and the center of the eye $x_t$ is set equal to the center of the iris $x_c$ so that the iris drags the eye-template toward the eye.

2.–3. The coefficients of the intensity and edge forces for the circle are increased. This helps scale the circle to the correct size of the iris. After this stage the position and size of the iris are considered essentially fixed and their derivatives are weighted by a small constant (see table 2) that destroys the symmetry in the $E_{\text{prior}}$, ensuring that the parameter values of the iris can influence the parameter values of the remainder of the template, but not vice-versa.

4. The template interacts with just the peak field, so that is parabolic boundaries rotate and translate to the right location.

*Table 1.* Change of energy coefficients with time. †indicates epochs using all points within $R_b$. Other weighting schedules are possible.

| Epoch | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $k_1$ | $k_2$ | $k_3$ | $k_4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1† | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2† | 100 | 100 | 0 | 0 | 400 | 0 | 0 | 0 | 0 | 0 |
| 3† | 200 | 250 | 0 | 0 | 300 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 200 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 50 | 200 | 0 | 100 | 1 | 5 | 5 | 0.01 |
| 6 | 200 | 400 | 160 | 400 | 150 | 100 | 1 | 5 | 5 | 0.01 |
| 7 | 200 | 500 | 200 | 400 | 125 | 100 | 1 | 5 | 5 | 0.01 |
| 8 | 200 | 250 | 100 | 400 | 63 | 100 | 1 | 5 | 5 | 0.01 |

*Table 2.* How parameters are updated: Each table entry has the form $K_{ep}\hat{p}_{ep}$ where $K$ is a constant (usually 0 or 1) and $\hat{p}$ is a parameter. During each descent step of epoch $e$, each parameter $p$ is updated according to $(p_{\text{new}} - p_{\text{old}}) = -(\delta t)K_{ep}\, \partial E/\partial \hat{p}_{ep}$. Other update schedules are possible.

| *Epoch* | $x_e$ | $y_e$ | $a$ | $b$ | $c$ | $\theta$ | $x_c$ | $y_c$ | $r$ | $p_1$ | $p_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $x_c$ | $y_c$ | 0 | 0 | 0 | 0 | $x_c$ | $y_c$ | 0 | 0 | 0 |
| 2–3 | $4x_c$ | $4y_c$ | $r$ | $r$ | $r$ | 0 | $4x_c$ | $4y_c$ | $r$ | $r$ | $-r$ |
| 4 | $x_e$ | $y_e$ | 0 | 0 | 0 | $.001\theta$ | 0 | 0 | 0 | $p_1$ | $p_2$ |
| 5 | $x_e$ | $y_e$ | 0 | $b$ | 0 | $.01\theta$ | 0 | 0 | 0 | $p_1$ | $p_2$ |
| 6–8 | $x_e$ | $y_e$ | $a$ | $b$ | $c$ | 0 | $.01x_c$ | $.01y_c$ | $.01r$ | $p_1$ | $p_2$ |

*Table 3.* Tolerances $\epsilon(e, p)$ determining convergence with respect to parameter $p$ during epoch $e$. Convergence occurs when $\Delta_p = |p_{\text{old}} - p_{\text{new}}| < \epsilon(e, p)\ \forall\ p$ for $M$ consecutive descent steps. $dt$ is set at each step so that max over $p$ of $\Delta_p = \Delta_{\max}$. A blank space indicates the parameter is not tested.

| Epoch | $x_e$ | $y_e$ | $a$ | $b$ | $c$ | $\theta$ | $x_c$ | $y_c$ | $r$ | $p_1$ | $p_2$ | $\Delta_{\max}$ | $M$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | 1.5 | 1.5 | | | | .20 | 100 |
| 2 | | | | | | | 1.5 | 1.5 | .3 | | | .10 | 100 |
| 3 | | | | | | | .5 | .5 | .3 | | | .10 | 50 |
| 4 | | | | | | .2 | | | | .3 | .3 | .10 | 100 |
| 5 | | | | | | .1 | | | | .2 | .2 | .10 | 100 |
| 6–7 | | | .2 | .2 | .2 | | | | | | | .10 | 100 |
| 8 | .2 | .2 | .2 | .2 | .2 | .2 | .2 | .2 | .2 | .2 | .2 | .02 | 100 |

5. The position of the white boundaries are fine tuned by adding interactions between the edge and intensity fields.
6.–8. The position of the template are fine tuned by allowing all fields to interact and by allowing all parameters to change.

The program changes epoch automatically when it has reached a steady state of the energy function with the appropriate coefficient values (i.e., when it thinks it has accomplished its goals for that epoch). See table 3 for details.

Figure 3 illustrates the program running in the different epochs. Note that the template can start some distance away from the eye, can scale the iris, rotate the eye and lock onto the edges. Figure 4 shows the final state of the system on several different eyes. The runtime for the program is between five and ten minutes on a SUN4.

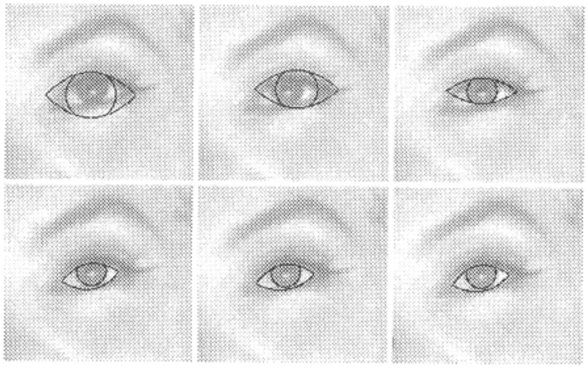

*Fig. 3.* Eye template at different times during the minimization. A dynamic sequence for the eye left to right and top to bottom. The first frame shows the initial configuration and the remaining frames show the results at the ends of the epochs (the final frame combines the results of the fifth and sixth epoch). Note how the valley force pulls the template in, the intensity helps to scale the circle correctly, the peaks orient it and the edges and the intensity fine tunes it.

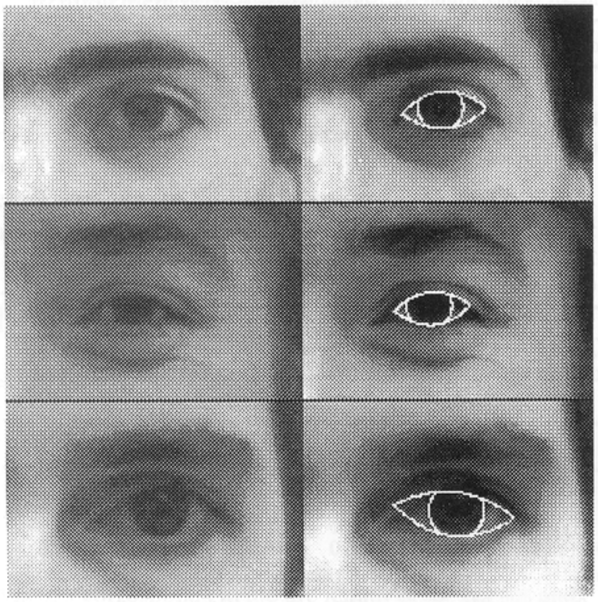

*Fig. 4.* Final results of the eye templates. The right column shows the final state of the template acting on the eyes in the left column. Note that a small error occurs in the alignment of the bottom template. This is due to a strong intensity peak on the eyelid and some shadow in the eye.

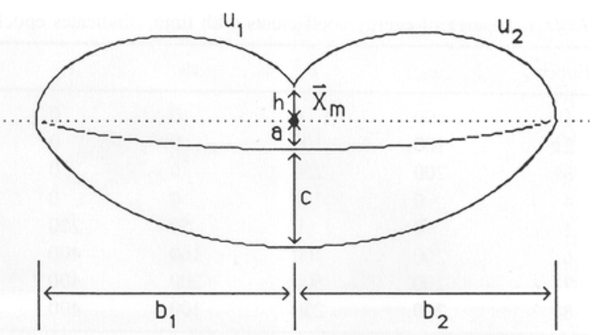

*Fig. 5.* The mouth-closed template. The mouth is centered on a point $\vec{x}_m$ and has an orientation $\theta$. The left and right boundaries are at distances $b_1$ and $b_2$ from $\vec{x}_m$. The intersection of the upper two parabolas occur directly above $\vec{x}_m$ at a height of $h$. The lower two parabolas have maximum distances from the central line (shown dotted) of $a$ and $a + c$.

## 5 The Mouth Template

The appearance of the mouth varies considerably depending on whether it is open or closed (i.e., on whether the teeth are visible or not). We define a mouth-closed template and a mouth-open template, although, as we will demonstrate later, the mouth-open template is also capable of detecting closed mouths.

For a closed mouth the most salient feature is a deep valley in the image intensity where the lips meet. There will also be edges at the top and bottom of the lips although the latter edge is often very weak. This motivates the mouth-closed template of figure 5.

We define the mouth-closed template in terms of a coordinate system $(x, y)$ centered on a point $\vec{x}$(the center of the mouth) and inclined at an angle $\theta$ (the orientation of the mouth). The positive $y$ direction points downward for consistency with the coordinate system used on the computer screen. The template is defined as follows: (1) The edge at the top of the upper lip is represented,by two parabolas $P_u$ which intersect above the center of the mouth. These are given by the lines

$$y_{u1}(x) = \left\{\frac{h - a + u_1 b_1}{b_1^2}\right\} x^2 + u_1 x + (a - h)$$

$$y_{u2}(x) = \left\{\frac{h - a - u_2 b_2}{b_2^2}\right\} x^2 + u_2 x + (a - h) \qquad (13)$$

The coefficients $u_1$ and $u_2$ help characterize the upper two curves but have no simple geometric interpretation. The coefficients $a$, $b_1$, $h$, $b_2$, $c$ are defined in figure 5.

(2) The edge at the bottom of the lower lip is represented by a parabola $P_l$

$$y_l(x) = (a + c)\left\{1 - \frac{4}{(b_1 + b_2)^2}\left[x - \frac{b_2 - b_1}{2}\right]^2\right\} \qquad (14)$$

(3) The valley at the intersection of the lips is represented by a parabola $P_v$

$$y_v(x) = a\left\{1 - \frac{4}{(b_1 + b_2)^2}\left[x - \frac{b_2 - b_1}{2}\right]^2\right\} \qquad (15)$$

The template depends on 10 parameters $\vec{g} = (a, b_1, b_2, u_1, u_2, h, c, \vec{x}, \theta)$ and its potential energy function $E_{M-C}(\vec{g})$ is given by

$$E_{M-C} = E_v + E_e + E_u + E_b + E_{ch} + E_p \qquad (16)$$

where the valley potential, calculated along the valley parabola $y_4(x)$, is

$$E_v = \frac{-\,c_1}{|P_v|}\int_{P_v} \Phi_v(\vec{x})\, ds \tag{17}$$

the edge potential, calculated along the upper lip parabolae $y_1(x)$ and $y_2(x)$ and the lower-lip parabola $y_3(x)$, is

$$E_e = \frac{-c_2}{|P_u|}\int_{P_u} \Phi_e(\vec{x})\, ds - \frac{c_3}{|P_l|}\int_{P_l} \Phi_e(\vec{x})\, ds \tag{18}$$

and the internal potentials are

$$E_u = \frac{k_1}{2}\,(u_1 + u_2)^2$$

$$E_b = \frac{k_2}{2}\,(b_1 - b_2)^2$$

$$E_{ch} = \frac{k_3}{2}\,(c - \lambda h)^2$$

$$E_p = \frac{k_4}{2}\left(\frac{h}{b_1 + b_2}\right)^2 \tag{19}$$

The internal potentials attempt to make the top two parabolas similar (the $E_u$ term), place the center of the mouth midway between the corners ($E_b$ term), make the width of the lower lip roughly $\lambda$ times that of the upper lip ($E_{ch}$ term) and encourage the top lip to bend down at the center (the $E_p$ term, this helps prevent the upper lip getting pulled up to the nose).

Typical values for the constants are $c_1 \approx 1000$, $c_2 \approx 100$, $c_3 \approx 15$, $\lambda \approx 2.0$, $k_1 \approx 0.1$, $k_2 \approx 1.0$, $k_3 \approx 0.1$, $k_4 \approx 1000.0$. The update is again done by steepest descent.

An additional expansion force attempts to make the mouth as wide as possible based on the average strength of the valley force on the center parabola. This gives additional update energy terms to $b_1$ and $b_2$ of

$$k_7\,\frac{1}{|P_v|}\int_{P_v} \Phi_v(\vec{x})\, ds \tag{20}$$

For an open mouth, the most salient features are the teeth and the two intensity valleys separating them from the upper and lower lips. It is tempting to describe the teeth as intensity peaks. However, although the teeth are strongly salient to human observers, they are often less bright than the specularities on the lips. There is, however, a strong edge field corresponding to the edges between the teeth. We therefore define a region corresponding to the teeth which is attracted to both peaks and edges. See figure 6.

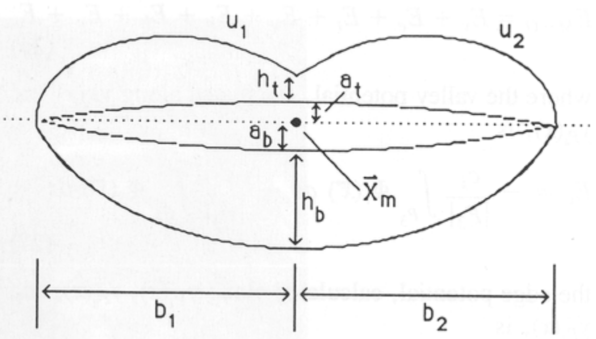

*Fig. 6.* The mouth-open template. The mouth is centered on a point $\vec{x}_m$ and has an orientation of $\theta$. The left and right boundaries are at distances $b_1$ and $b_2$ from $\vec{x}_m$. The intersection of the upper two parabolas occurs directly above $\vec{x}_m$ at a height of $a_t - h_t$. The central two parabolas have maximum distances from the central line (shown dotted) of $a_t$ and $a_b$. The bottom parabola has maximum distance $h_b + a_b$.

The mouth-open template can be obtained by having two valley parabolas instead of one. The region between them corresponds to the teeth. The parabolas for the top of the upper lip ($P_1$ and $P_2$), bottom of the upper lip ($P_3$), top of the lower lip ($P_4$), and bottom of the lower lip ($P_5$) are given by equations (21), (22), (23), (24), and (25), respectively (as before we use a coordinate system centered on $\vec{x}$, oriented by an angle $\theta$ and with the positive $y$ axis pointing downward)

$$y_{P_1}(x) = \left\{\frac{h_t - a_t + u_1 b_1}{b_1^2}\right\} x^2 + u_1 x + (a_t - h_t) \tag{21}$$

$$y_{P_2}(x) = \left\{\frac{h_t - a_t - u_2 b_2}{b_2^2}\right\} x^2 + u_2 x + (a_t - h_t) \tag{22}$$

$$y_{P_3}(x) = a_t\left\{1 - \frac{4}{(b_1 + b_2)^2}\left[x - \frac{b_2 - b_1}{2}\right]^2\right\} \tag{23}$$

$$y_{P_4}(x) = a_b\left\{1 - \frac{4}{(b_1 + b_2)^2}\left[x - \frac{b_2 - b_1}{2}\right]^2\right\} \tag{24}$$

$$y_{P_5}(x) = (a_b + h_b)\left\{1 - \frac{4}{(b_1 + b_2)^2}\left[x - \frac{b_2 - b_1}{2}\right]^2\right\} \tag{25}$$

The region between the upper and lower lips has area $(2/3)(b_1 + b_2)(a_b - a_t)$. The template depends on 11 parameters $\vec{g} = (a_t, a_b, b_1, b_2, u_1, u_2, h_t, h_b, \vec{x}, \theta)$ and its potential energy function $E_{M-O}(\vec{g})$ is given by

$$E_{M-O} = E_v + E_e + E_t + E_u + E_b + E_h + E_p + E_a \tag{26}$$

where the valley potential, calculated along $y_{P_3}(x)$ and $y_{P_4}(x)$, is

$$E_v = -\frac{c_1}{|P_3|}\int_{P_3} \Phi_v(\vec{x})\, ds - \frac{c_2}{|P_4|}\int_{P_4} \Phi_v(\vec{x})\, ds \tag{27}$$

the edge potential, calculated along $y_{P_1}(x)$, $y_{P_2}(x)$, and $y_{P_5}(x)$, is

$$E_e = -\frac{c_3}{|P_1|}\int_{P_1} \Phi_e(\vec{x})\, ds - \frac{c_3}{|P_2|}\int_{P2} \Phi_e(\vec{x})\, ds - \frac{c_4}{|P_5|}\int_{P_5} \Phi_e(\vec{x})\, ds \tag{28}$$

the teeth potential, which attempts to maximize the average intensity and strength of edges in the teeth region $R_1$ between the upper and lower lips $y_{P_3}(x)$ and $y_{P_4}(x)$, is

$$E_t = -\frac{c_5}{|R_1|}\int_{R_1} \{\Phi_p(\vec{x}) + \lambda\Phi_e(\vec{x})\}\, dA \tag{29}$$

and the internal potentials are

$$\begin{aligned} E_u &= \frac{k_1}{2}(u_1 + u_2)^2 \\ E_b &= \frac{k_2}{2}(b_1 - b_2)^2 \\ E_h &= \frac{k_3}{2}(h_b - \lambda h_t)^2 \\ E_p &= \frac{k_4}{2}\left(\frac{h}{b_1 + b_2}\right)^2 \\ E_a &= k_5\,|a_b - a_t| \end{aligned} \tag{30}$$

The first four internal potentials are the same as for the mouth-open template (allowing for changes in notation). The new potential $E_a$ ensures that the mouth is clamped shut in the absence of teeth. Note that this degenerates to the mouth-closed template when $a_b = a_t$. The constant $\lambda$ is usually set to be 1.0.

The dynamics follow by steepest descent in parameter space as before.

An additional force $F_r$ is defined to help open the mouth. It is based on the total strength of the peak and edge forces in the entire teeth region. This gives additional energy update terms for $a_t$ and $a_b$ of

$$-k_6\frac{1}{|R_1|}\int_{R_1} \{\Phi_p(\vec{x}) + \lambda\Phi_e(\vec{x})\}\, dA \tag{31}$$

and

$$+k_6\frac{1}{|R_1|}\int_{R_1} \{\Phi_p(\vec{x}) + \lambda\Phi_e(\vec{x})\}\, dA \tag{32}$$

respectively.

## 6 Simulation Results for Mouths

Again the potential fields of the valleys, peaks and edges are computed before the program starts. Figure 7 shows the typical form of these fields.

The system worked well for both the mouth-closed and mouth-open templates after some preliminary experimentation to fix the values for the coefficients.

For the mouth-closed template there were two epochs: (i) Coefficients are high for the valley forces and zero for the edge forces. The valley term pulls the template to the mouth, scales and orients it. The expansion force also helps to scale it. (ii) The edge coefficients are increased. The edges help adjust the positions of the edge boundaries.

Figure 8 demonstrates this method and shows the time history of a simulation. Figure 9 shows the final positions of the template on several images.

For the mouth-open template there were again two epochs: (i) Coefficients are high for the valley forces and the teeth forces. They are zero for the edge forces on the boundary. The teeth forces pull the template to the mouth, scale and orient it. (ii) The edge coefficients are increased. The edge forces help adjust the positions of the edge boundaries. Stage (ii) has not yet been implemented for this case.

Figure 10 shows the potential fields for open mouth for valleys, peaks, and edges. Note how a combination of the peak and edge fields is needed to specify the teeth. Figure 11 shows dynamic sequences of the mouth-open template successfully running on both open and closed mouths. For the closed mouth the area of the teeth region of the template shrank to zero.

The run time on a SUN4 was again between five and ten minutes.

## 7 Tracking

It is straightforward to adapt the deformable templates for tracking. Here we describe a straightforward implementation that tracks eyes automatically given an initial position and a set of potential fields. For the first

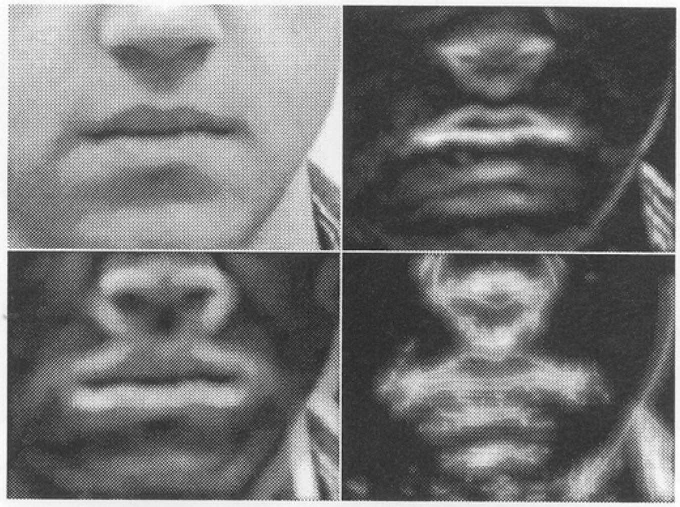

*Fig. 7.* The potential fields for a closed mouth: (a) the original mouth, (b) the valley field, (c) the peak field, and (d) the edge field. The figures are organized left to right and top to bottom. The strengths of the fields are shown in greytone; white is strong and black is weak.

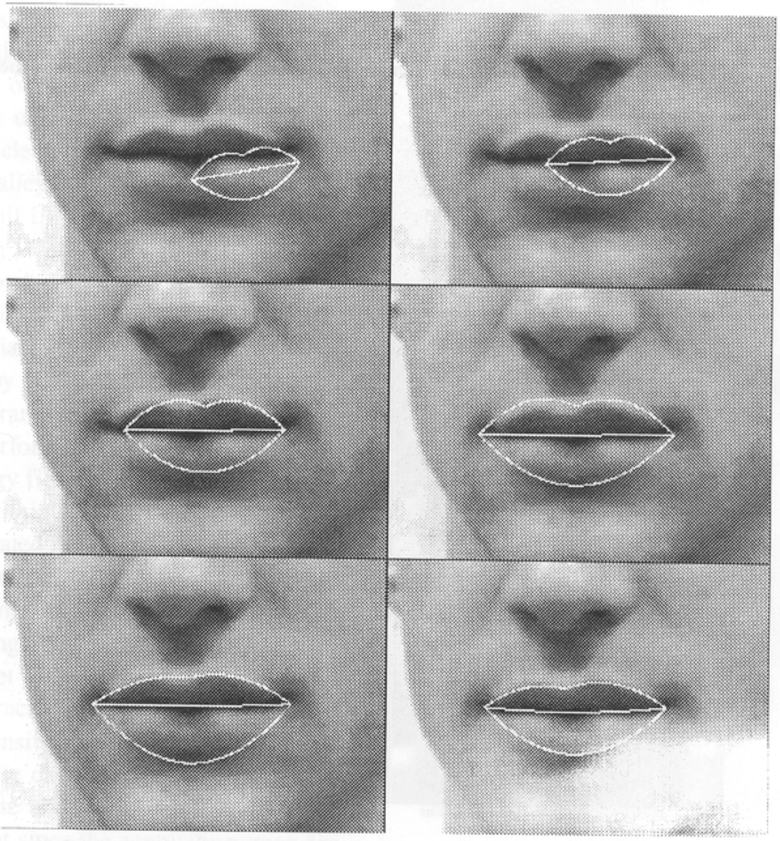

*Fig. 8.* A dynamic sequence for mouth-closed left to right and top to bottom. Observe that the valley pulls the template in, scales it and orients it. The edge forces do fine tuning.

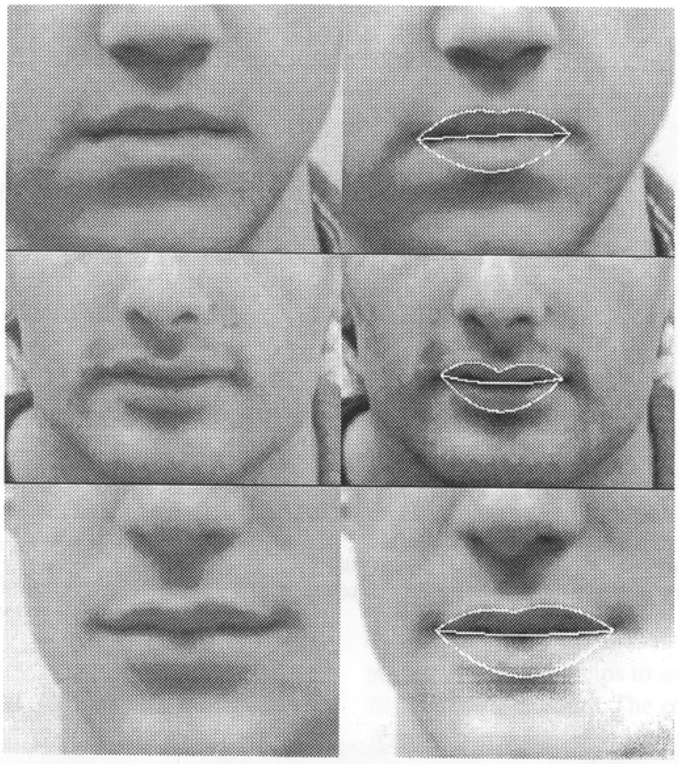

*Fig. 9.* Final results for the mouth-closed templates. The right column shows the final state of the template acting on the mouth in the left column.

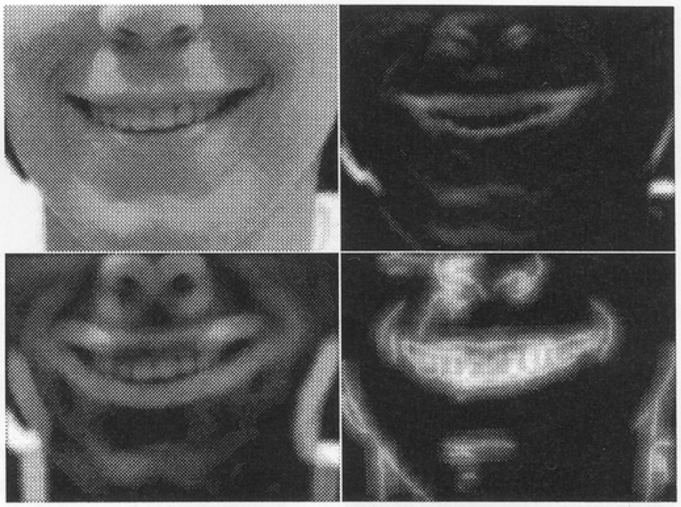

*Fig. 10.* The potential fields for an open mouth: (a) the original mouth, (b) the valley field, (c) the peak field, and (d) the edge field. Note the strong combination of peaks and edges in the teeth region. The figures are organized left to right and top to bottom. The strengths of the fields are shown in greytone, white is strong and black is weak.

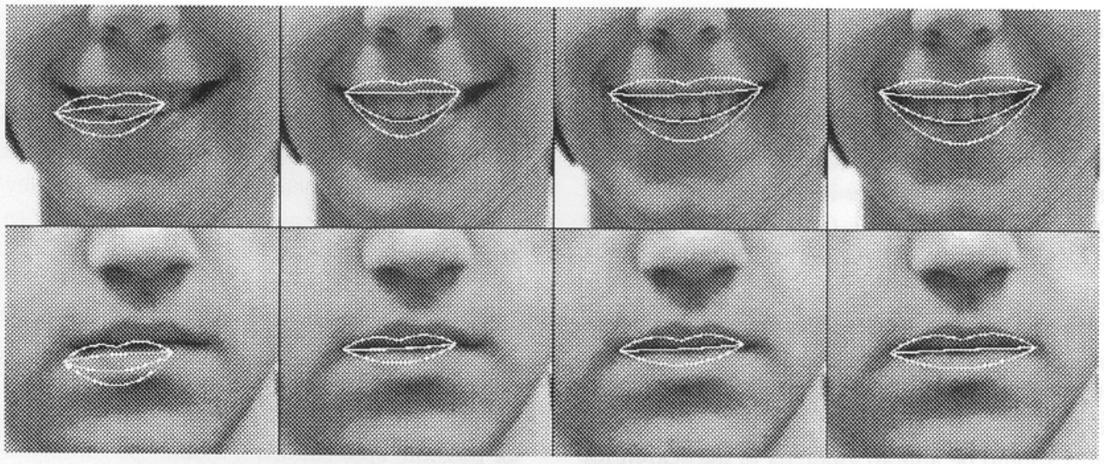

*Fig. 11.* A dynamic sequence for the mouth-open template on an open and a closed mouth. In the upper picture the template is pulled in by the peak and edge forces from the teeth. The two valley parabolas are pulled to the bottom of the upper lip and the top of the lower lip. In the lower picture the template is mainly pulled in by the valley forces and the two valley parabolas contract to the same curve. Note that the edge fields were switched off for these simulations, so the final positions of the edges should be ignored.

frame, we initialize the template by hand as before, but for succeeding frames, we use as the initial position the best fit of the preceding frame. This method obviously succeeds only as long as the best fit at time $t - 1$ lies in the basin of attraction of the system at time $t$. In turn this will be true only if: (i) the deformations and movements of the eye are small (e.g., on the order of the diameter of the iris) and (ii) the potential fields are sufficiently clean and accurate (e.g., eyebrows are not marked as valleys and the smoothing scale $\rho$ is long enough to pull in templates from far away).

The first criterion is met by using a high enough frame rate. The second criterion is met by constructing potential fields in the following way: (i) choose a scale for the potential fields; (ii) construct peak and valley energy fields by running robust peak and valley templates over each frame (see Hallinan 91; Yuille and Hallinan 92); (iii) perform nonmaximum suppression on the resulting energy fields; and (iv) suppress peaks that do not appear within the surround of any valley. Edge fields are generated by thresholding the gradient magnitude. The results are smoothed as before with an exponential decay kernel. However, because valleys and peaks appear as single points in the results, the smoothing scale is set longer for valleys and shorter for peaks to improve tracking by valleys and to minimize conflicts between intensity peaks on the skin and those on the whites. To these clean potential fields we then apply the eye template above.

A final point is that since the acquisition stage provides a close fit to the eye and since the deformation is not expected to be great from frame to frame, we can economize on computation by relaxing the convergence criteria and by turning on the valley-radius momentum term in the first epoch.

Results of this system for a real eye are shown in figure 12. Note that a completely automatic system could be built by incorporating the automatic acquisition described by Hallinan (1991). Also, robust potential fields could be generated on the fly using as a scale estimate the radius of the iris in the previous frame.

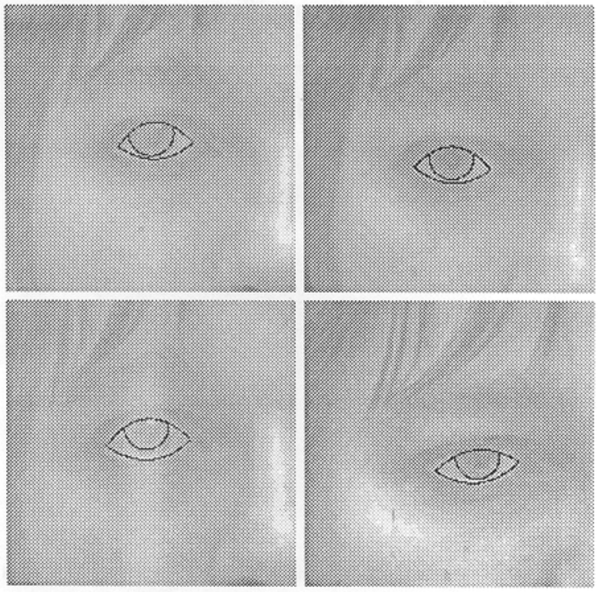

*Fig. 12.* A sequence of images of eyes being tracked. Both the subject's head orientation and direction of gaze are varying.

## 8 Extensions and Future Work

It seems relatively straightforward to find templates for the other "internal" features of the face, such as eyebrows, noses, chins, and moustaches (indeed work on detection of face outlines is reported by Brunelli (1991) and Bennet and Craw (1991)). It is less clear how to generalize this idea to find "external" features such as the ears or hair, or to find internal regions such as the forehead or the cheeks. However, Identikit programs used by police forces are able to represent a large variety of faces by using a comparatively small number of templates (120 eyes, for example). Such programs should be able to guide us in the search for reliable ways to parameterize features. Our strategy for the implementation was to use preprocessing to set the initial values of the template parameters. An alternative method would be to start several deformable templates in parallel and see which gives the best results. This would require some criteria for selecting the best fit. A natural choice would be the one with the lowest final energy function. This, however, might need to be supplemented by taking into account the spatial relationships to other features and the a priori probability of the final parameter values. In some special cases it may be possible for the energy to be low but for the parameter values to be extremely unlikely. Such a situation can occur if the mouth templates gets started on the eye and becomes grotesquely deformed—see figure 13.

Interactions between templates may also be necessary for detection. The features of the face are constrained to have certain spatial relationships with each other, and this should affect the detection. These forces might be mediated by springs. Moreover, once a feature is detected the potential fields corresponding to it can be removed, thereby making it easier to detect the remaining features. For example, once the eyebrows are detected, removing the valley fields associated with them would make it easier to detect the eyes.

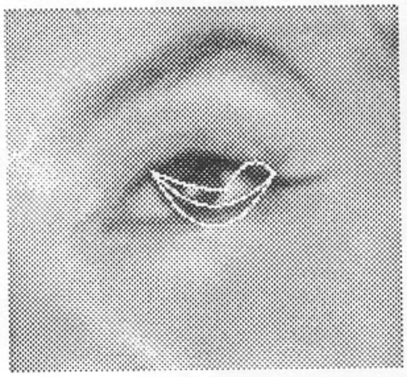

*Fig. 13.* A grotesquely deformed template. The mouth-closed template is pulled into the eye. Although the resulting energy can be small the final parameter values of the template are so strange that it cannot be interpreted as a mouth.

Progress on these problems is reported by Hallinan (1991). This includes making the matching criterion robust and automating the selection of valleys and initial template positions.

Deformable templates seem to have a large number of possible applications. Another possibility is to use them for perceptual grouping; a set of these templates (capable of describing many salient shapes) could interact with the image and those with the best matches (least energy) would be chosen to order the image. The visual system would "hypothesize" many different structures, allow them to interact with the image, and then choose the best. It is unclear, however, how many templates would be needed, how many different starting points in the image, and how computationally intensive this procedure would be. A second possibility is to use deformable templates to describe the three-dimensional surfaces and allow the reflectance function to be specified by a finite number of parameters (allowing for possible directions of the light source, different types of reflectance, etc.). Suppose, for example, that we have a deformable template representing the three-dimensional geometry of the nose. The reflectance function might also be specified by a finite set of parameters (allowing Lambertian plus specularity). The geometry of the nose and its reflectance will then be described by a finite set of parameters. This can then be related to the image by the image-irradiance equation. There will then (usually) be a sufficient number of equations to solve for the parameters.

## 9 A General Formulation of Deformable Templates

Finally we describe a more general formulation of deformable templates (Grenander et al. 1991; Yuille and Hallinan 1992; Hallinan 1991). The deformable template consists of three basic elements:

1. A parameterized *Geometrical Model* for the feature including prior probabilities for the parameters. This corresponds to a geometric measure of fitness.

2. An *Imaging Model* to specify how a deformable template of specific geometry will give rise to specific intensities in the image. This can be expressed as an imaging measure of fitness.
3. An algorithm using the geometrical and imaging measures of fitness to match the template to the image.

It is attractive to formalize this definition in terms of probabilities. Suppose $T(\vec{g})$ specifies the geometrical model of the template with prior probability $P(\vec{g})$ on the template parameters $\vec{g}$. The imaging mode $P(I \mid T(\vec{g}))$ gives the probability of producing an image $I$ from a template $T(\vec{g})$. Thus $P[I \mid T(\vec{g})]P(\vec{g})$ can be used to synthesize features.

Bayes' theorem can be used to obtain a measure of fitness. We write

$$P[T(\vec{g}) \mid I] = \frac{P[I \mid T(\vec{g})]P[T(\vec{g})]}{P(I)} \tag{33}$$

This gives us a probability of detection of a template, $P[T(\vec{g}) \mid I]$, in terms of the imaging model and the prior probabilities. By maximizing $P[T(\vec{g}) \mid I]$ with respect to $\vec{g}$ we can find locally optimal candidate matches. One can also use $P[T(\vec{g}) \mid I]$ as a confidence criterion for the matches.

## 10 Conclusion

A serious problem for detection of edges, or other feature, seems to lie in combining local information, which may be easily obtained, into a global structure.

For the purpose of detecting facial features, however, a lot more a priori information is available and a deformable template is able to capture it. Moreover, such templates are not only able to detect a feature but can also provide a description of it for classification and matching to a data base.

## Acknowledgments

We would like to thank the Brown, Harvard, and M.I.T. Center for Intelligent Control Systems for a United States Army Research Office grant number DAAL03-86-C-0171, the National Science Foundation for grant IRI-9002141, and Roger Brockett for his support. Conversations with Jim Clark, David Mumford, Petros Maragos, Mark Nitzberg, Gaile Gordon, and Roger Brockett were extremely useful.

## References

Ayache, N., Boissonnat, J.D., Brunet, E., Cohen, L., Chieze, J.P., Geiger, B., Monga, O., Rocchisani, J.M., and Sander, P. 1989. Building highly structured volume representations in 3D medical images, *Comp. Assoc. Rad., Proc. Intern. Symp.*, Berlin, pp. 765–772.

Ballard, D.H., and Brown, C.M., 1982. *Computer Vision*. Prentice-Hall: Englewood Cliffs, NJ.

Bennett, A., and Craw, I., 1991. Finding image features for deformable templates and detailed prior statistical knowledge. *Preprint. Dept. Mathematical Sciences, University of Aberdeen. Scotland.*

Brunelli, R., 1991. Face recognition: Dynamic programming for the detection of face outline, Technical Report 910-06, Instituto per la Ricerca Scientifica e Tecnologica, Trento, Italy.

Burr, D.J., 1981a. A dynamic model for image registration, *Comput. Graph. Image Process.* 15:102–112.

Burr, D.J., 1981b. Elastic matching of line drawings,, *IEEE Trans. Patt. Anal. Mach. Intell.* 3(6):708–713.

Durbin, R., and Willshaw, D.J., 1987. An analogue approach to the travelling salesman problem using an elastic net method, *Nature*, pp. 689–691.

Durbin, R., Szeliski, R., and Yuille, A.L., 1989. An analysis of the elastic net approach to the travelling salesman problem, *Neural Computat.* 1:348–358.

Grenander, U., Chow, Y., and Keenan, D.M., 1991. *HANDS. A Pattern Theoretical Study of Biological Shapes*. Springer-Verlag: New York.

Hallinan, P.W., 1991. Recognizing human eyes. *Proc. Conf. 1570, SPIE* San Diego.

Huber, P.J., 1981. *Robust Statistics*. Wiley: New York.

Kanade, T., 1977. Computer recognition of human faces, *Birkhauser Verlag*, Basel and Stuttgart.

Kass, M., Witkin, A., and Terzopoulos, D., 1987. Snakes: Active contour models, *Proc. 1st Intern. Conf. Comput. Vis.* London, June.

Lipson, P., Yuille, A.L., O'Keefe, D., Cavanaugh, J., Taaffe, J. and Rosenthal, D., 1990. Deformable templates for feature extraction from medical images. *Proc. 1st Europ. Conf. Comput. Vis.* Antibes, France.

Maragos, P., 1987. Tutorial on advances in morphological image processing and analysis, *Optical Engineering* 26:623–632, July.

Parisi, G., 1988. *Statistical Field Theory*, Addison-Wesley: Reading, MA.

Pentland, A., 1987. Recognition by parts. *Proc. 1st Intern. Conf. Comput. Vis.*, London, June.

Serra, J., 1982. *Image, Analysis and Mathematical Morphology*, Academic Press: New York.

Shackleton, M.A., and Welsh, W.J. 1991. Classification of facial features for recognition *Proc. Comput. Vis. Patt. Recog.* Hawaii, pp. 573–579.

Yuille, A.L., Cohen, D.S., and Hallinan, P.W., 1989. Feature extraction from faces using deformable templates, *Proc. Comput. Vis. Patt. Recog.*, pp. 104–109.

Yuille, A.L., and Hallinan, P.W., 1992. Deformable templates, in *Active Vision*, ed. A. Blake and A.L. Yuille. M.I.T. Press: Cambridge, to appear.