FACIAL FEATURES LOCALIZATION IN FRONT VIEW HEAD AND SHOULDERS IMAGES

Adnan M. Alattar
Digimarc
One Centerpointe Drive, Suite 500
Lake Oswego, OR 97035-8615
e-mail: aattar@gte.net

Sarah A. Rajala
College of Engineering, Box 7904
North Carolina State University
Raleigh, NC 27695-7904
e-mail: sar@eos.ncsu.edu

ABSTRACT

The computerized process of locating human facial features such as the eyes, nose and mouth in a head and shoulders image is crucial to such applications as automatic face identification and model-based video coding. In this paper, a new model-based algorithm for locating these major features is developed. The algorithm estimates the parameters of the ellipse which best fits the head view in the image and uses these parameters to calculate the estimated locations of the facial features. It then refines the estimated coordinates of the eyes, mouth, and nose by exploiting the vertical and horizontal projections of the pixels in windows around the estimated locations of the features. The algorithm has been implemented and tested with over twelve hundred images, and simulation results indicate that the algorithm is robust to variations in subject head shape, eye shape, age, and motion such as tilting and nodding of the head.

1. INTRODUCTION

The computerized process of locating human facial features such as the eyes, nose, and mouth in a head and shoulders image is very important to such applications as automatic face recognition [5] and model-based video coding [2]. It is also essential to the low-level primitive extraction process needed for the emerging MPEG-7 Multimedia Content Description Interface Standard [10].

Several approaches have been proposed in the literature for locating facial features in front view head and shoulders images. Sakai, et al. [14] use the vertical and the horizontal signatures of the pixels in a moving slit of fixed dimensions to iteratively determine the locations of the facial features. Horizontal and vertical signatures are also used by Brunelli and Poggio [3] to locate facial features in a feature-based face recognition system.

Yuille, et al. [15] use deformable templates to search for the facial features around the peaks and valleys of the intensity image. A similar approach is used by Hallinan [7] to detect the eyes in an image. Both Yuille and Hallinan report that deformable templates are very sensitive to the starting location of the template. Moreover, adjusting the templates and minimizing the associated energy function is computationally intensive. Huang and Chen [9] use simple heuristics to obtain rough contours for the facial features; the contours are subsequently refined using deformable templates or active contours.

Reisfed and Yeshurun [13] exploit facial symmetry to locate the eyes and the nose. This approach is computationally intensive due to the required search for symmetry points. Chibelushi, et al. [6] use coarse-to-fine fuzzy reasoning and multi-resolution spatio-temporal information to locate the eyes and the mouth from a sequence of head and shoulders images. Preliminary results are successful but sensitive to noise, eyes closure, and spectacles. Ahmed Fadzil and Abu Baker [1] use a multi-layered neural network to search the head area and locate the eyes. Chang and Huang [4] and Ohya, et. al. [12] use skin color to locate facial features in colored images. Nugroho et. al. [11] use genetic algorithm to locate the facial features.

In this paper, a new model-based algorithm is developed which quickly and accurately locates the eyes, nose, and mouth in front-view head and shoulders images. Section (2) describes the estimation process of the locations of the facial features in terms of the parameters of a head model. Section (3) describes the refinement process of these locations. Section (4) discusses the implementation of the facial feature location algorithm and presents simulation results and an analysis. Section (5) contains the conclusions.

2. ESTIMATING THE LOCATIONS OF THE EYES, NOSE AND MOUTH

The head of a human being can be thought of as an ellipsoid that sits on the top of the torso and is attached to it at the center by the neck. The head movement can then be viewed as rotations α, β, and θ around the *X*, *Y*, and *Z*-axes, respectively, which pass through the center of the ellipsoid. Under this ellipsoid model, the head view in a front-view image has the shape of an ellipse, and the

angles α and β are both zero. The tilt angle θ of the head is indicated by the orientation of the major axis of the ellipse (see Fig. (1a)).

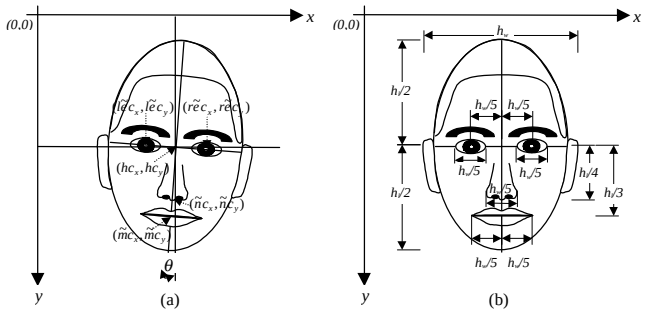

Fig. 1. Head Model for Head and Shoulders Image; (a) elliptical head model (b) head geometry.

If *h(x,y)* is the function that describes the head contour in a front-view head and shoulders image, then it can be approximated under this model by the ellipse function *f(x,y)*.

$$h(x,y) \approx f(x,y) = ax^2 + 2bxy + cy^2 + dx + ey - 1 = 0 \quad (1)$$

where *a*, *b*, *c*, *d*, and *e* are the five parameters of the ellipse. Once the parameters (a,b,c,d,e) of the ellipse are calculated, the head center (hc_x, hc_y), the head tilt angle θ, the head width h_w, and the head length h_l can be calculated.

It is well known from the art of drawing the human head that the average head is approximately five eye-lengths wide [8] (see Fig. (1b)). Both eyes lie on the line midway between the top of the head and the bottom of the chin, and the distance between them is approximately equal to one eye-length. The nose starts at the center of the face and descends to a point mid-way between the center of the face and the base of the chin. The width of the nose at the base is also equal to approximately one eye-length. The mouth barrel starts at the base of the nose and extends two thirds of the distance down from the nose to the chin. The distance between the central line of the mouth and the center of the face is approximately one third of the head length. The corners of the mouth align with the centers of the eye sockets. Therefore, the length e_l of each eye is

$$e_l = \frac{h_w}{5} \quad (2a)$$

and the estimated center $(\tilde{re}c_x, \tilde{re}c_y)$ of the right eye is

$$\begin{aligned} \tilde{re}c_x &= hc_x + e_l \cdot \cos\theta \\ \tilde{re}c_y &= hc_y + e_l \cdot \sin\theta \end{aligned} \quad (2b)$$

Similarly, the estimated center $(\tilde{le}c_x, \tilde{le}c_y)$ of the left eye:

$$\begin{aligned} \tilde{le}c_x &= hc_x - e_l \cdot \cos\theta \\ \tilde{le}c_y &= hc_y - e_l \cdot \sin\theta \end{aligned} \quad (2c)$$

The nose length n_h and the nose width n_w (at the base) are

$$(n_h, n_w) = \left(\frac{h_l}{4}, \frac{h_w}{5}\right) \quad (3a)$$

The estimated center $(\tilde{nt}c_x, \tilde{nt}c_y)$ of the nose tip is

$$\begin{aligned} \tilde{nt}c_x &= hc_x - \frac{h_l}{2}\sin\theta \\ \tilde{nt}c_y &= hc_y + \frac{h_l}{2}\cos\theta \end{aligned} \quad (3b)$$

The height m_h and the width m_w of the mouth barrel are

$$(m_h, m_w) = \left(\frac{h_l}{6}, \frac{2h_w}{5}\right) \quad (4a)$$

The estimated center $(\tilde{mc}_x, \tilde{mc}_y)$ of the mouth is

$$\begin{aligned} \tilde{mc}_x &= hc_x - \frac{h_l}{3}\sin\theta \\ \tilde{mc}_y &= hc_y + \frac{h_l}{3}\cos\theta \end{aligned} \quad (4b)$$

3. REFINING THE LOCATIONS OF THE EYES, NOSE AND MOUTH

If *f(x,y)* is a pixel value in a given *w*x*h* tilted rectangular window, where *x* and *y* are relative to the window's own rectangular coordinate system that is aligned with its sides, then define the *x*-projection, *h(x)*, and *y*-projection, *v(x)*, of the window as follows

$$\begin{aligned} h(x) &= \sum_{y=0}^{h} f(x,y) \\ v(y) &= \sum_{x=0}^{w} f(x,y) \end{aligned} \quad (5)$$

In order to find the exact locations of the eyes, nose, and mouth, the *x* and/or the *y* projections in tilted rectangular windows over the estimated locations of the features are used. The tilt angle of the window is the same as the head tilt angle θ. The size of the eye window is $3.5el \times eh$, and it is centered on the center of the ellipse (*cx*, *cy*). The size of the nose window is $nw \times nh$, and it is centered at ($\tilde{nt}c_x$, $\tilde{nt}c_y$ *-nh/4*). The size of the mouth window is $mw \times mh$, and it is centered at ($\tilde{mc}_x$, $\tilde{mc}_y$).

Fig. (2a) and Fig. (2b) show *v(y)* and *h(x)*, respectively, for the eyes window of a typical front-view head and shoulders image. In both figures, the solid curve represents the actual projection, while the dotted curve represents the projection thresholded at the average value of the projection curve. The valleys in Fig. (2a) and Fig. (2b) occur at the eyes' sockets, and therefore indicate the vertical and horizontal coordinates of the two eyes. Hence, these coordinates are used to refine the initial estimates of the eyes' locations.

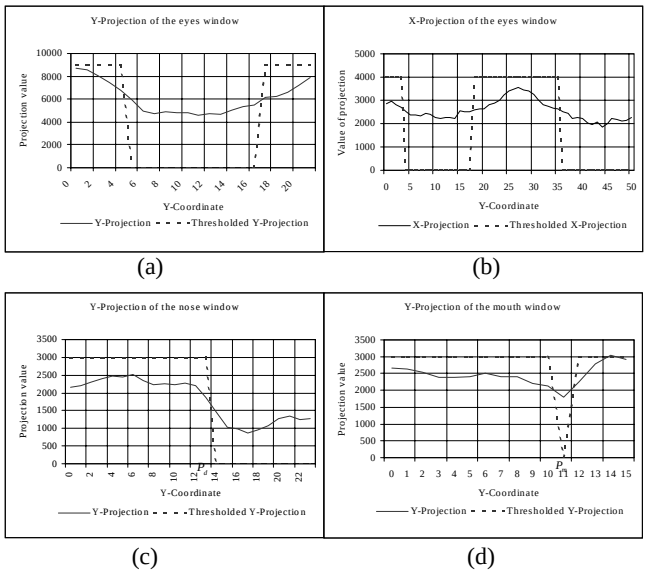

Fig. 2. Projections of the image data in the tilted windows around the facial features; (a) Y-Projection of the eyes window, (b) X-Projection of the eyes window, (c) Y-Projection of the nose window, (d) Y-Projection of the mouth window.

Let the distance between the center of the valley in Fig. (2a) and the center of the eye's window be m. Also, let the distance between the center of the left valley in Fig. (2b) and the center of the eye's window be k, and the distance between the center of the right valley in Fig. (2b) and the center of the eye's window be l. Then the refined centers of the left eye (lec_x, lec_y) and the right eye (rec_x, rec_y) expressed in terms of m, k, l, hc_x, hc_y, and θ are

$$\begin{aligned} lec_x &= hc_x - k\cdot\cos\theta + m\cdot\sin\theta \\ lec_y &= hc_y - k\cdot\sin\theta - m\cdot\cos\theta \end{aligned} \tag{6a}$$

$$\begin{aligned} rec_x &= hc_x + l\cdot\cos\theta + m\cdot\sin\theta \\ rec_y &= hc_y + l\cdot\sin\theta - m\cdot\cos\theta \end{aligned} \tag{6b}$$

Fig. (2c) shows the y-projection $v(y)$ for the nose window of a typical front-view head and shoulders image. The solid curve represents the actual projection, while the dotted curve represents $v(y)$ thresholded at the average value of $v(y)$. The tip of the nose corresponds to the discontinuity point P_d of the thresholded curve. If the distance between P_d and the center of the nose window is $\left(n-(ntc_y-\frac{nh}{4}-hc_y)\right)$, then the distance between P_d and the center of the head is n. Therefore, the refined center of the tip of the nose (ntc_x,ntc_y) is

$$\begin{aligned} ntc_x &= hc_x - n\cdot\sin\theta \\ ntc_y &= hc_y + n\cdot\cos\theta \end{aligned} \tag{7}$$

Fig. (2d) shows the y-projection $v(y)$ for the mouth window of a typical front-view head and shoulders image. The solid curve represents the actual projection, while the dotted curve indicates the location of the minimum point of $v(y)$. The middle line of the mouth corresponds to the minimum point at P_m. If the distance between P_m and the center of the mouth window is $\left((mc_y-hc_y)-o\right)$, then the distance between P_m and the center of the head is o. Therefore the refined center of the mouth (mc_x,mc_y) in terms of o, hc_x, hc_y, and θ can be easily determined to be

$$\begin{aligned} mc_x &= hc_x - o\cdot\sin\theta \\ mc_y &= hc_y + o\cdot\cos\theta \end{aligned} \tag{8}$$

4. SIMULATION RESULTS AND ANALYSIS

The developed Facial Feature Location (FFL) algorithm has been successfully implemented and tested with 1,285 images taken from fifteen video sequences of talking heads. These sequences contain various subjects and a broad spectrum of facial characteristics including shouting; a lively, chatting child; squinting; eyeglasses; tilting and turning head; nodding; side-to-side rocking motion; smiling; rolling of the eyes; and laughing.

Fig. (3) shows intermediate and final results of detecting facial features for an image taken from one of the fifteen test sequences. The algorithmically determined locations of the facial features are marked by crosshairs on the original image as shown in Fig. (3d).

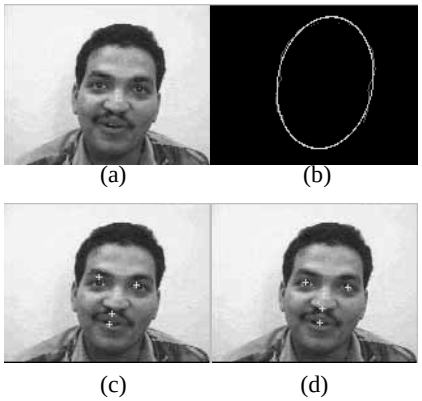

Fig. 3. Intermediate and final results of detecting facial features; (a) Original Image, (b) The ellipse that approximates the head contour (c) Approximate locations of the facial features (d) Refined locations of the facial features.

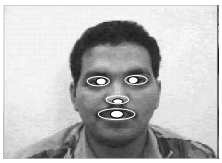

Fig. 4. Fourbull's-Eyes template used to subjectively test FFL algorithm's accuracy.

To test the FFL algorithm's accuracy in locating the eyes, nose, and mouth, the 1,285 crosshaired output images have

been subjectively tested using a "fourbull's-eyes" template as shown in Fig. (4). For a given output image, an algorithmically determined facial feature location is classified as a *hit* (*H*) if the point of intersection of the crosshairs is subjectively determined to be within the solid ellipse. It is classified as a *near hit* (*NH*) if the point of intersection of the crosshairs is determined to be outside of the solid ellipse, but inside of the larger ellipse. It is classified as a *miss* (*M*) if the point of intersection of the crosshairs is determined to be outside of both of the ellipses over the respective facial feature.

Results from the fifteen test sequences indicate that the FFL algorithm has excellent performance with nearly 92% hit, 7% near hit and less than 1% miss. The percentage of hit for the eyes or the nose is over 97%, but it is about 76% for the mouth. The near hit percentage for the mouth is 23%. The low performance of the algorithm with the mouth is mainly due to the confusion with the beard or mustache.

Results also indicate that the algorithm is robust to changes in subject face shape, eye shape, age, and head motion such as nodding. Even in images where degradation is seen in locating the mouth (for example, when the subject has a moustache), the algorithm gives excellent performance in locating the eyes and nose. Since the algorithm was explicitly designed for full-frontal view, head and shoulders images, the algorithm's performance is slightly degraded when the head in an input image rotates away from the full-frontal view. Even with difficult cases (such as images with moustaches or head rotation) a percentage of hits of nearly 92% has been obtained for the overall algorithm performance.

5. CONCLUSIONS

A new model-based algorithm has been developed which successfully locates the major features of the face in a front-view head and shoulders image. The algorithm is based on modeling the head in the image by an ellipse. The algorithm has been successfully implemented, and tested. Simulation results indicate that the algorithm is robust to variations in subject head shape, eye shape, age, and motion such as tilting and nodding of the head. The algorithm has an overall performance of 92%. This performance can be further enhanced by including more characteristics (e.g. color) in the refinement process of the location of the mouth. This will help to distinguish the mouth from the moustache and the beard.

REFERENCES

[1] M. H. Ahmed Fadzil and H Abu Bakar, "Human face recognition using neural networks," in *IEEE Int. Conf. Image Proc.*, vol. 3, pp. 936-938, 1994.

[2] K. Aizawa and T. S. Huang, "Model-based image coding: advanced video coding techniques for very low bit-rate applications," *Proc. of the IEEE*, vol. 83, no. 2, pp. 259-271, Feb. 1995.

[3] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. on PAMI*, vol. 15, no. 10, pp. 1042-1052, Oct. 1993.

[4] T. C. Chang and T. S. Huang, "Facial feature extraction from color images," *12th IAPR Int. Conf. Pattern Recognition*, vol. 2, pp. 39-43, October 1994.

[5] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.

[6] C. C. Chibelushi, F. Deravi, and J. S. Mason, "Face segmentation using fuzzy reasoning," in *Proc. IEEE 1st Int. Conf. Image Proc.*, vol. 3, pp. 518-522, 1994.

[7] P. Hallinan, "Recognizing human eyes," in *SPIE Proc.: Geometric Methods in Computer Vision*, vol. 1570, pp. 214-226, 1991.

[8] B. Hogarth, *Drawing the Human Head.* 1st ed., New York: Watson-Guptill, 1965.

[9] C. Huang and C. Chen, "Human facial feature extraction for face interpretation and recognition," *Pattern Recognition*, vol. 25, no. 12, pp. 1435-1444, 1992.

[10] MPEG Requirements Group, "MPEG-7: context and objectives," *Doc. ISO/MPEG N1733*, MPEG Stockholm Meeting, July 1997.

[11] H. Nugroho, S. Takashashi, Y. Ooi, and S. Ozawa, "Detecting human face from monocular image sequences by genetic algorithms," *ICASSP'97*, pp. 2533-2535, April 1997.

[12] J. Ohya, Y. Kitamura, F. Kishino, and N. Terashima, "Virtual space teleconferencing: real-time reproduction of 3d human images," *Visual Commun. and Image Representation*, vol. 6, no. 1, pp. 1-25, March 1995.

[13] D. Reisfed and Y. Yeshuran, "Robust detection of facial features by generalized symmetry," in *Proc. 11th Int. Conf. on Pat. Recog.*, pp. 117-120, 1992.

[14] T. Sakai, M. Nagao, and Takeo Kanade, "Computer analysis and classification of images of human faces," in *Proc.First USA-Japan Computer Conf.*, pp. 55-62, 1972.

[15] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Proc. IEEE Computer Soc. Conf. Computer Vision and Pat. Recog.*, pp. 104-109, 1989.