Face Detection and Facial Feature Extraction
Using Color, Shape and Symmetry-Based Cost Functions

Eli Saber
Xerox Corporation
435 W. Commercial St.
East Rochester, NY 14445
saber@roch803.mc.xerox.com

A. Murat Tekalp
Dept. of Electrical Engineering
University of Rochester
Rochester, NY 14627
tekalp@ee.rochester.edu

Abstract

This paper describes an algorithm for detecting human faces and subsequently localizing the eyes, nose, and mouth. First, we locate the face based on color and shape information. To this effect, a supervised pixel-based color classifier is used to mark all pixels which are within a prespecified distance of “skin color.” This color-classification map is then subject to smoothing employing either morphological operations or filtering using a Gibbs random field model. The eigenvalues and eigenvectors computed from the spatial covariance matrix are utilized to fit an ellipse to the skin region under analysis. The Hausdorff distance is employed as a means for comparison, yielding a measure of proximity between the shape of the region and the ellipse model. Then, we introduce symmetry-based cost functions to locate the center of the eyes, tip of nose, and center of mouth within the facial segmentation mask. The cost functions are designed to take advantage of the inherent symmetries associated with facial patterns. We demonstrate the performance of our algorithm on a variety of images.

1. Introduction

Automatic detection and recognition of faces and facial expressions from still images and video is an emerging application. A complete facial image processing system should be able to: 1) localize faces in a given image, 2) identify and pin-point facial features, 3) recognize people, 4) describe facial expressions, and 5) locate faces based on a textual description. The majority of the research efforts have been focused on the recognition aspect with the assumption that the location of the face is known a priori. This assumption is well suited for certain specific applications. However, in general, the face must be localized before any recognition can be performed. This paper proposes a system to detect human face based on color and shape information, and then locate the eyes, nose, and mouth through the use of symmetry-based cost functions.

The progress in face detection and recognition is summarized in [3]. In [6], Craw et al. described a multiresolution scheme for extracting the head and facial features from mug-shot images. Turk et al. [14] proposed a Karhunen-Loeve (KL) based approach for locating faces by utilizing “eigenfaces.” Use of the KL transform has also been proposed in [9] where the inherent symmetry of the face is incorporated in an “eigenpicture” representation to improve the localization. However, the localization performance decreases quickly with changes in scale and orientation. Deformable template models have been utilized in [16, 7] to locate the eyes and mouth. In addition, active contour models were employed in [8] to capture the eyebrows, nostrils and face. These techniques rely heavily on “near” perfect segmentations or edge detection. Furthermore, the extracted contours are highly dependent on the initialization of the snake or active contour model, and on the parameters involved in defining these models. In [15], Yang et al. proposed a hierarchical three level knowledge-based system for locating human faces and facial features. Chang et al. [2] proposed a color segmentation and thresholding based algorithm to pinpoint the eyes, nostrils and mouth in color “head and shoulder” images. The skin segmentation is performed on a pixel-by-pixel basis, where a pixel is classified as skin if its chromaticity falls within a certain region of the chromaticity space. The eyes, nostrils, and mouth are located within pre-determined bounding boxes within the skin mask by thresholding the low intensity pixels, the normalized red component, and the normalized red + blue – 2* green component respectively. Sung et al. [13] described an approach for face detection by utilizing 19 × 19 window patterns and exhaustively searching a given image at all possible scales. Colmenarez et al. [5] introduced an algorithm for detecting faces based on a symmetry measurement that is designed

1015-4651/96 $5.00 © 1996 IEEE
Proceedings of ICPR '96

654
The contribution of this paper is to describe an algorithm for detecting human faces using color and shape, and then localizing the eyes, nose, and mouth through the use of symmetry-based cost functions. The algorithm consists of: 1) supervised skin/non-skin color classification, 2) shape classification and 3) eye, nose, and mouth localization. The skin/non-skin classification is performed by utilizing the chrominance channels of the YES color space [12] followed by morphological or Gibbs random field model-based smoothing. The shape classification is achieved by employing the eigenvalues/eigenvectors of the skin spatial covariance matrix. Finally, the eyes, nose, and mouth centers are localized by utilizing cost functions designed to take advantage of the inherent symmetries associated with human faces.

2. The Method

The overall algorithm is shown in Figure 1. First, an adaptive color classification algorithm is applied to the E and S channels to segment the image into two classes: skin and “other”. This is followed by morphological or Gibbs random field model-based smoothing to yield contiguous regions. Then, the eigenvalues and eigenvectors of the skin spatial covariance matrix are utilized to fit an ellipse to the skin region. The Hausdorff distance is employed as a means for comparison, yielding a measure of proximity between the shape of the region and the ellipse model. Finally, symmetry-based cost functions are introduced to locate the eyes, tip of nose, and center of mouth within the facial segmentation mask. The details are described below.

2.1. Color space

It is desirable to use a color space which is robust against shading, highlights, etc. Examples of these spaces include the YIQ and YCrCb “luminance-chrominance” spaces. However, it is also generally agreed that there does not exist a single color space which is good for all images [10]. In this paper, we use the YES space [1] given by:

\[
\begin{bmatrix}
Y \\
E \\
S
\end{bmatrix} = 
\begin{bmatrix}
0.253 & 0.684 & 0.063 \\
0.500 & -0.500 & 0.000 \\
0.250 & 0.250 & -0.500
\end{bmatrix} 
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

The YES space has been chosen because: 1) it is computationally efficient (the E and S channels can be computed from the R, G, and B by shifting bits rather than multiplication), and 2) it is free of singularities (nonlinear spaces may have singularities). However, the approaches and algorithms developed in the following sections can be utilized with any color space.

2.2. Supervised Skin/Non-Skin Color Classification

The supervised pixel-based color classification module implements the algorithm presented in [12]. In summary, the class-conditional pdf of the chrominance components \(w_{ij} = [E_{ij} \ S_{ij}]^T\) belonging to the skin class \(x_{ij}\) is modeled by a two-dimensional Gaussian, where the mean vector and the covariance matrix are estimated from an appropriate training set. Then, a binary hypothesis test has been employed to decide whether each pixel in a given image belongs to the skin class or not. The thresholds can be estimated either at run time from user specified confidence bounds, or pre-computed by using receiver operating characteristic analysis on a set of training images. The main advantage of this approach is its computational efficiency. Our results indicate that color can serve as a powerful initial classifier for locating faces.

2.3. Post-processing for Smoothing

A choice of smoothing via morphological operations or GRF filtering has been added following the pixel-based color classification to obtain contiguous label clusters. The morphological procedure consists of three erosions followed by three dilations using a $3 \times 3$ kernel. The GRF algorithm is the same as the one employed in the unsupervised segmentation [11], except it is initialized with the result of pixel-based color classification, and executed for a few iterations. In summary, an iteration is composed of three steps: Starting with an initial segmentation, the first step encompasses the calculation of the means for all pixels belonging to a particular class within a window, whose size is initially taken to be equal to the size of the image itself. In the second step, the segmentation labels are updated by utilizing the iterated conditional modes (ICM) algorithm. This procedure often converges to a local minimum of the Gibbs potential within a relatively small number of cycles. The last step consists of eliminating “small” classes, which is accomplished by combining them with their nearest color neighbors. Upon convergence, the size of the window, employed in the computation of the means, is then reduced by a factor of two in both horizontal and vertical directions, and the procedure is repeated until the window size is smaller than a user specified threshold. As a result, the algorithm starts with global estimates and slowly adapts to the local characteristics of each region.
2.4. Shape Classification

This step is utilized as a shape classifier in order to eliminate those regions that are “dissimilar” to a face template, whose shape is modeled in terms of an ellipse. Let $m_1$ and $m_2$ denote the horizontal and vertical axis respectively. Using the spatial coordinates of the skin identified pixels within a contiguous region, we form the spatial covariance matrix

$$R = \begin{bmatrix} \sigma_{m1}^2 & \sigma_{m1m2} \\ \sigma_{m1m2} & \sigma_{m2}^2 \end{bmatrix}$$

whose elements refer to average distances along the respective coordinates of the skin pixels from the shape centroid. The eigenvalues of $R$, $\lambda_1$, $\lambda_2$, provide us with a reasonable estimate of the spread of the skin region in the direction of the corresponding eigenvectors $(v_1, v_2)$. The directions of the eigenvectors indicate the principal axis of the skin classified region, as shown in Figure 2, and thereby utilized as the major and minor axis of the ellipse shape model

$$\frac{m_1^2}{\lambda_1} + \frac{m_2^2}{\lambda_2} = c$$

where $c$ represents the “radius” of the ellipse. Its center is taken as the centroid of the skin region as shown in Figure 2. We utilize the Hausdorff distance as a means for comparison between the skin region shape and the ellipse model for various values of $c$ in the interval $[c_{min}, c_{max}]$, where $c_{min}$ is set to 1. $c_{max}$ is taken as twice the value of $c$ where all the pixel coordinates of the skin region are enclosed within the ellipse border as shown in Figure 2. Given two finite point sets $S1 = \{e_0, e_1, \ldots, e_p\}$ and $S2 = \{f_0, f_1, \ldots, f_q\}$, the Hausdorff distance is defined as:

$$H(S1, S2) = \max(h(S1, S2), h(S2, S1))$$

where:

$$h(S1, S2) = \max_{e \in S1} \min_{f \in S2} \|e - f\|$$

and $\|\cdot\|$ is the Euclidean norm. The Hausdorff distance is employed because: 1) it follows from Eq. (3) that if the computed Hausdorff distance is $d$, then every point in $S1$ must be within a distance $d$ of some point in $S2$ and vice versa, 2) no explicit pairing of points between $S1$ and $S2$ is required, 3) the number of points in $S1$ does not have to be equal to that of $S2$, as would be the case if one were to utilize a mean square error measurement, and 4) its ease of computation. Consequently, the value of $c$ that minimizes the Hausdorff distance is chosen as the optimum value. The corresponding distance value is utilized as a measure of similarity between the shape of the skin region and the ellipse model, rejecting those that result in a measure greater than a pre-specified threshold.

2.5. Symmetry-based Cost Functions for Eyes, Nose, and Mouth Localization

Once the facial pattern has been detected and the major and minor axis identified, we introduce symmetry-based cost functions to locate the eyes, nose, and mouth within the facial segmentation mask. Our goal is to: 1) find the two “holes” within the mask that are most likely to have occurred due to the location of the eyes, and 2) locate the nose and the mouth once the eyes have been identified.

The “eyes” cost functions are designed to take advantage of the following facts on an upright “mug-shot” facial pattern: 1) eyes are located on a line which is parallel to the minor axis represented by the direction of the eigenvector corresponding to the smaller eigenvalue; 2) eyes are symmetric with respect to the major axis represented by the direction of the eigenvector corresponding to the larger eigenvalue; 3) eyes are equidistant from the the minor axis; 4) eyes are, for the most part, the closest “holes” in the skin segmentation mask to the minor axis; and 5) eyes are located above the minor axis.

Let $(u_n, v_n), n = 0, 1, \ldots, l-1$ define the centroids of the “holes” within the skin segmentation mask, and $(U_c, V_c)$ indicate the centroid of the mask itself. The first cost function is defined as:

$$C^1_n = \text{Abs} \left[ \frac{v_n - v_{n-1}}{u_n - u_{n-1}} \right]$$

By examining Eq. (5), it can be easily seen that $C^1_n$ reaches its minimum when the two centroids are located horizontally from each other; and its value increases as they deviate from the horizontal, reaching a maximum when they are located in a vertical position.

The second cost function is designed to take advantage of the fact that the eyes are generally symmetric with respect to the major axis represented by the direction of the eigenvector corresponding to the larger eigenvalue. It is defined as:

$$C^2_n = \text{Abs} \left[ \text{Abs}(u_n - U_c) - \text{Abs}(u_{n-1} - U_c) \right]$$

A quick examination of Eq. (6) indicates that $C^2_n$ reaches its minimum when the centroids of the two “holes” in question are located symmetrically from each other across the vertical axis represented by the eigenvector corresponding to the largest eigenvalue.

Our third proposed cost function is defined as:

$$C^3_n = \text{Abs} \left[ \text{Abs}(v_n - V_c) - \text{Abs}(v_{n-1} - V_c) \right]$$

$C^3_n$ reaches its minimum if the two “holes” under examination are equidistant from the horizontal minor axis represented by the direction of the eigenvector corresponding to the smaller eigenvalue.
The fourth cost function is designed to locate the closest "holes" in the skin segmentation mask to the horizontal minor axis. It is defined as:

$$C_n^4 = \text{Abs}(v_n - V_c) + \text{Abs}(v_{n-1} - V_c)$$  \hspace{1cm} (8)$$

Its minimum is reached when the centroids of the two "holes" are located exactly on the minor axis; and its value increases as these centroids are moved further away.

The fifth and final cost function is defined as:

$$C_n^5 = (v_n - V_c) + (v_{n-1} - V_c)$$  \hspace{1cm} (9)$$

$C_n^5$ is negative when the two "holes" are located above the horizontal minor axis; and positive otherwise.

The weighted combination of these cost functions is defined as:

$$C_n = \sum_{r=1}^{5} w_r C_n^r$$  \hspace{1cm} (10)$$

Its minimum represents the "two holes" within the facial segmentation mask that are most likely to have occurred due to the eyes. The centroid of these holes is utilized to mark the location of the eyes.

Upon identifying the eyes, we utilize the distance $d$ between the eye centers, as shown in Figure 3, to locate the tip of the nose and the center of the mouth. The thresholds $t_1$ and $t_2$ are determined experimentally from a database of faces.

3. Results

The test images utilized are of RGB type, where each channel is quantized to 8 bits/pixel. Figures 4a and 5a demonstrate the performance of our algorithm where the center of the eyes, the tip of the nose, and the center of the mouth are indicated by a " + " superimposed on the original image. It can be easily seen that the facial features were located accurately for all practical purposes. Figures 4b and 5b show the ellipse fit to the skin classified region superimposed on the GRF based skin classification.

4. Conclusions

This paper described an algorithm for detecting faces based on color and shape, and locating the eyes, nose and mouth through the use of symmetry-based cost functions. The shape classification and eyes, nose, and mouth localization are limited to frontal facial views, and consequently would not yield meaningful results on profiles. The detection of the tip of the nose and center of the mouth is dependent on the location of the center of the eyes within the skin segmentation mask. The proposed approach can be utilized as a pre-processing step to a face recognition system.

References

Figure 1. Face detection and eye, nose and mouth localization algorithm

Figure 2. Ellipse fit to skin region

Figure 3. Nose and mouth location

Figure 4. Face detection and eyes, nose, and mouth localization

Figure 5. Face detection and eyes, nose, and mouth localization