POSE INVARIANT FACE DETECTION [1]

Manoj Seshadrinathan and Jezekiel Ben-Arie

ECE Dept., M/C 154, University Of Illinois at Chicago,
851 S.Morgan St., Chicago, IL 60607, USA
Phone/Fax:001-312-996-2648, E-mail:benarie@ece.uic.edu

Abstract: *In this paper, we present a novel method for pose invariant face detection in color images. The novelty in our method of face detection arises from the integration of evidence from various independent sources such as color, frequency response and geometric shape information. Skin color is detected in the Y-Cb-Cr color space using a RPROP neural network. The shape information is derived from a novel ellipse area criterion and then it is formulated to compute a probabilistic score of the connected components that represent the faces in the images. The third source is based on Gabor 2D filters that are used to obtain frequency signature of faces in the images. The final likelihood of a face is a combination of the individual probabilities of color, shape and Gabor response.*

Key Words: - *Face detection, Skin color, Y-Cb-Cr Color Space, RPROP neural network, Gabors, Ellipse Area Criterion, KL Transform*

1. INTRODUCTION

Face detection in images has a wide range of applications such as video surveillance, human computer interface and tracking. The original contribution and the main idea behind our work was integration of evidence from various independent information sources. Other works on face and head detection [6] are not robust and restrict themselves to identification of frontal views of faces and many base themselves on detection of facial features like eyes etc. This is a restriction that inhibits the performance of the face detection algorithms. Other works such as [7, 8] restrict themselves to recognition of individual faces whereas detection requires generic recognition. In contrast, our method is capable of detecting faces in a wide range of viewing directions.

As the first step in our algorithm, we detect skin areas in the image. Previous methods in skin color detection generate a skin color model [4, 3]. Based on the fact that skin color is clustered together in the Y-Cb-Cr color space we apply a direct neural network based approach to the skin color detection. We use a Resilient Propagation neural network (RPROP) [5] to separate skin and non-skin regions. The RPROP is a modified version of the Back-propagation algorithm. In the y-Cb-Cr model, the illumination is decoupled from the color information and hence makes our detection easier. In our work, we neglect the luminance since it is subject to changes and we just use the chrominance and the chromaticity values. We train the neural network on a wide variety of skin samples. Typical of this network is about three thousand skin samples and an equivalent amount of non-skin samples. False alarms were filtered out imposing an aspect ratio criterion of 1:3 on the connected components. For the shape information a probabilistic score is computed with respect to the area of the ellipse that could be formed with the major and minor axes of the connected components that represent the faces in the image. In the next step, hair regions are detected. Since it has several false alarms, this is used just to augment the skin score. Another information source

[1] This work was supported by the National Science Foundation (NSF) Grants No. IIS-0208300 and IIS-9979774.

for face detection is based on the Gabor 2D wavelets that are used to obtain the frequency signature of faces in the image. The responses of the Gabor wavelets closely resemble that of the human cortical cells. We train the Gabor wavelets on test images. We have a set of model signatures that are used to detect the location of faces in the image. The Gabor wavelets give a score for every pixel in the image. This score is a measure of the similarity that the pixel has to the center of a test face image. Assuming that the information sources we use are independent, the probability of a face is a combination of the individual probabilities of color, shape and Gabor response. The likelihood for a face center is computed by the average of these individual probabilities (shown in Section 5). Thus, by integrating the evidence a final score is obtained. Components with score above a threshold limit are classified as faces.

2. SKIN COLOR EXTRACTION

Human skin color, though it differs widely from person to person, is distributed over a very small area on the Cb-Cr plane [9],[2]. We use the Y-Cb-Cr color model for the detection of skin color. It is in the Y-Cb-Cr color space that the luminance information is decoupled from the color information. Furthermore, the Cb-Cr are the chrominance components used in MPEG and JPEG.

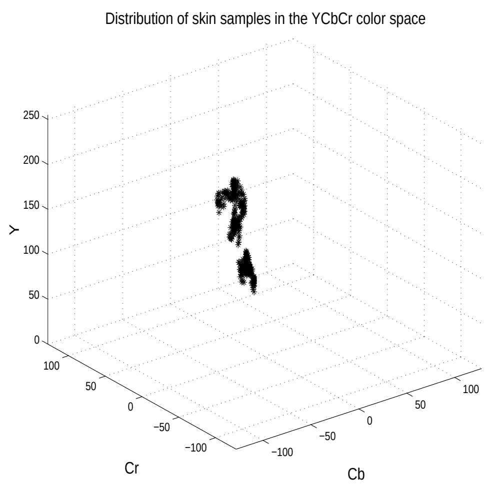


Fig. 1 Distribution in Y-Cb-Cr space

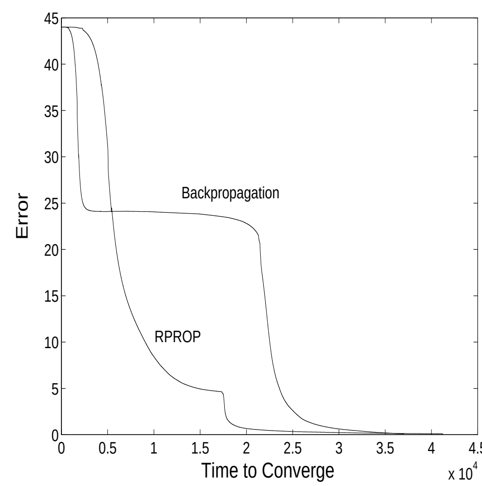


Fig. 2 Plot for Convergence

It is seen from Fig. 1 that the intensity value Y does not contribute much to finding skin color and that skin color forms a tight cluster in the Cb-Cr space. In [9] skin color classification was performed directly in Cb-Cr space without taking the Y value into account. Hence, we use the Cb-Cr values to train the RPROP neural network.

The Resilient Propagation (RPROP) Algorithm is a modified form of the backpropagation algorithm[5]. It has been established that for optimal convergence the hidden layer should have $2n+1$ perceptrons where n is the number of inputs. In our case, the number of inputs is two and hence 5 neurons would suffice. But, we increase the number of neurons in the hidden layers to make the network more flexible and learn faster. We also find that with two hidden layers the convergence of the network is faster and more smooth. So, in our case, we use a network with two hidden layers. The first hidden layer has 6 neurons and the second layer has 4 neurons. The output is of single dimension and has a value between 0 and 1.

The RPROP is an efficient new learning scheme, that performs a direct adaptation of the weight step based on local gradient information. We see from the graph in Fig. 2 that RPROP gave excellent convergence as opposed to traditional backpropagation.

3. ELLIPSE AREA CRITERION

The RPROP assigns a value [0,1] for each pixel. Each pixel in the image now has a value that denotes the possibility that this pixel belongs to the skin. The higher the value, the higher this possibility. A threshold value can then be established. Now, any pixel with value above threshold is denoted as 1 i.e. a skin pixel and all other pixels are denoted as 0. This forms a binary image. To find the face regions, the connected components in the image are labeled as face or no face.

Initial false alarms are filtered out by using the aspect ratio criterion of 1:3. We denote the aspect ratio by A_r. A_r is always less than 1 and is derived according to the equation below:

$$A_r = \begin{cases} \frac{W}{H} & \text{if } W \leq H \\ \frac{H}{W} & \text{if } W > H \end{cases}$$

where, W and H are the width and height. The aspect score S_a is 1 if $A_r > 0.33$ and 0 if $A_r \leq 0.33$. When $S_a = 0$ the region is eliminated. We assume that all face regions have approximately the shape of an ellipse. To find such an ellipse, we find for each connected component in the image the major and minor axes along its centroid. This is done by the KL transform. The coordinates of each of the connected components are subjected to the KL transform.The KL transform establishes a new co-ordinate system whose origin is at the centroid of the vector population and the axes are in the direction of the eigenvectors of the covariance matrix.

a. b.

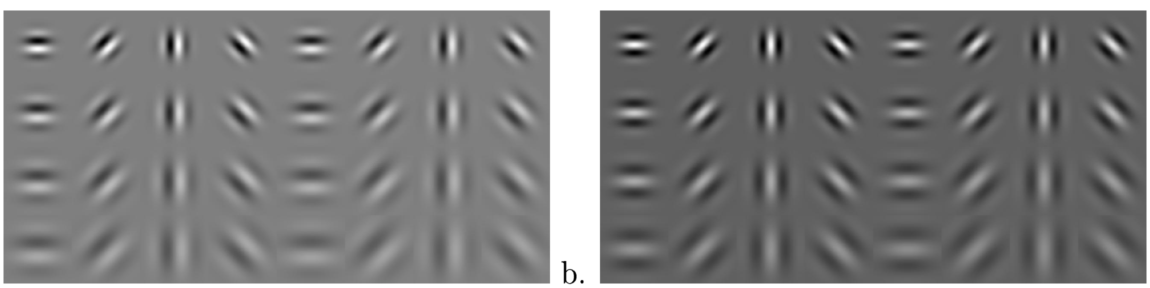


Fig. 3 The 64 Real and Imaginary Gabor filters used

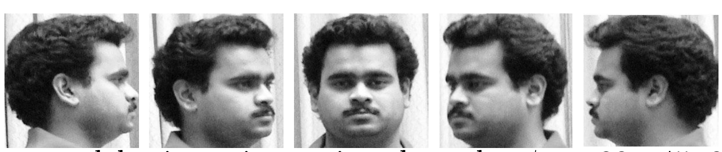


Fig. 4 Head training model orientations azimuth angles ψ = -90, -45, 0, +45, +90 degrees

After the transformed coordinates are obtained we have the major and minor axes of each component. The human face resembles an ellipse. Here we apply a simple but very effective criterion which we term the ellipse area criterion. In this we measure the similarity of each blob to an ellipse that could be formed by the major and minor axes of the connected component. Now, if the length of the major axis is l_1 and the minor axis is l_2, each connected component will have a score S_e based on its similarity to an ellipse expressed as

$$S_{ei}(x,y) = \frac{4 * N_i}{\pi * l_1 * l_2} \quad S_h(x,y) = \frac{4 * N_h}{\pi * l_1 * l_2} \tag{1}$$

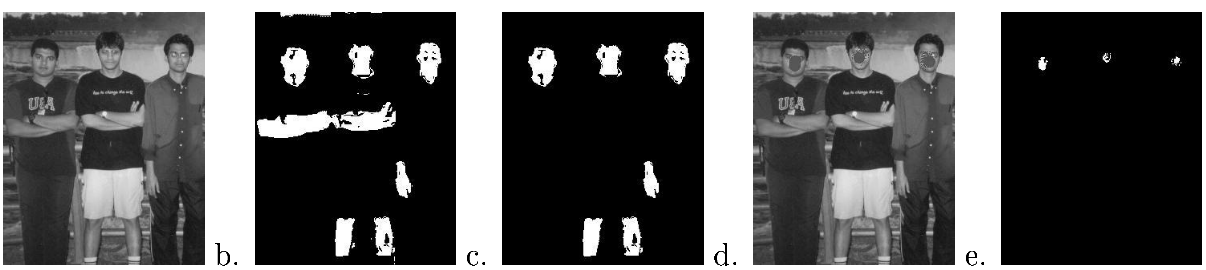
where i is the number of the connected component in the image and N_i is the number of pixels in the connected component i. Each connected component has a score based on the ellipse area criterion. This score is allotted to each and every pixel belonging to that connected component. The hair color is only used to augment the score based on the ellipse area criterion. This is because the RPROP based hair detection gives rise to many false alarms of dark backgrounds being detected as hair. Hence, we just consider as hair, only the hair colored region inside the ellipse formed by the major and minor axes of the skin blob. If the total number of hair pixels inside the region is N_h the score for hair S_h is obtained as shown in the equation above. $S_e(x,y)$ and $S_h(x,y)$ are actually one combined score, the sum of which is less than one.

4. FREQUENCY RESPONSE FROM GABOR 2D WAVELETS

The Gabor 2D wavelet is another independent evidence source which provides the frequency response characteristics of the image. The Gabor representation extended to two dimensions is,

$$\Phi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\ e^{\frac{-1}{2}[(\frac{x-x_0}{\sigma_x})^2+(\frac{y-y_0}{\sigma_y})^2]}\ e^{-j[u_0(x-x_0)+v_0(y-y_0)]} \tag{2}$$

where (x,y) is any point in the image and σ_x, σ_y denotes the scale of the Gaussian along the respective axes, (x_0, y_0) represents the center of the function in the spatial domain and (u_0, v_0) are the angular frequencies.



a. b. c. d. e.

Fig. 5(a) The Original Image, (b) detected skin regions, (c) regions passing the aspect ratio criterion, (d) detection by the Gabors (shown as dark blobs in the face), (e) Regions with high total scores $S_t(x,y)$

In our experiments we use sixty four gabors, at 4 orientations and 8 scales. The scaling factor has to be logarithmic and hence, σ ranges from 22 to 44 in multiples of 1.1040. It has been chosen so that the Gabors cover a frequency range of two octaves.

The volume of these Gabor filters is set to a constant to maintain a scale invariant amplitude of response when correlating the Gabor filters with an image. We use five training models to get the signatures of the Gabors for varying head pose. The Gabor projections on the five images are obtained. We have five sets of signatures one for each model. For extracting the signatures on these model images, the images are projected onto the Gabor filters. These yield five sets of 64 responses each. These signatures can be used to identify and detect faces in an image. To detect faces, the Gabor filters are correlated with the test image. Signatures are extracted for each pixel in the test image. We calculate the inner product of these signatures with each of the five model signatures and take the maximum value and assign it to the pixel. This way every pixel has a score $S_g(x, y)$ which is representative of its similarity to the model face.

5. INTEGRATING THE SCORES

Since it is assumed that the shape, color and frequency response information are statistically independent of each other, we obtain the combined score $S_t(x, y)$ for each pixel (x, y) by combining the three individual scores for each pixel which could be interpreted as proportional to the probability of classification. Hence,

$$S_t(x, y) = \left(\frac{S_h(x, y) + S_e(x, y) + S_s(x, y) + S_g(x, y)}{3}\right) * S_a \tag{3}$$

The second score $S_e(x, y)$ is representative of the shape information and is obtained by the ellipse area criterion. The score due to the presence of hair in the elliptical region S_h is just used to augment the skin score. The third score $S_g(x, y)$ is obtained from the projections of Gabor 2D wavelets.

6. EXPERIMENTAL RESULTS

In order to test the robustness of the face detection system several experiments were performed under different conditions. First, the skin regions of the test image are detected using skin color score $S_s(x, y)$. All pixels that have a score of $S_s(x, y) > t_s$ are selected where t_s is the threshold imposed. Fig. 5 (b) illustrates such thresholding of an image.

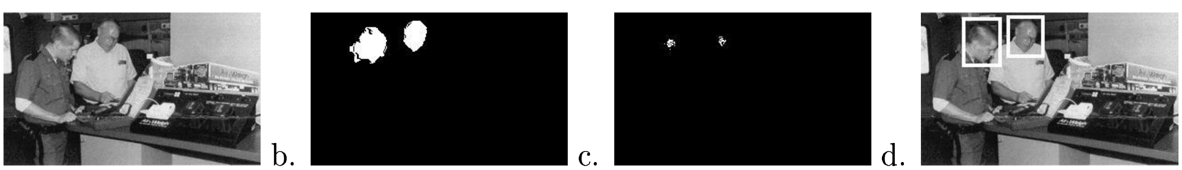

a. b. c. d.

Fig. 6 (a) Example of detection of faces in profile view with high clutter, (b) Regions passing the ellipse area criterion, (c) detection by the gabors and (d) detected faces are shown.

Initially, the entire CVL face database[1], which includes 114 images of persons is processed by our algorithm and all the 114 faces in these images are detected accurately using our methods. Our method works well for faces whose viewpoints vary between -90^o to $+90^o$ degrees of azimuth angle ψ with the frontal view defined as 0 degrees. The score $S_e(x, y)$ of the ellipse area criterion decreases when the absolute value of azimuth angle increases. A graph is plotted of the score of the ellipse area criterion $S_e(x, y)$ vs. the azimuth angle ψ in Fig. 7 (d). Detection results on random images on crowds are shown in Fig. 7 (a),(b) and (c). Here, we see that having a low threshold gives some false detections (Fig. 7 (b)). But using our threshold of 0.8 the false alarms are eliminated. In our tests, 482 faces out of 500 were detected giving a detection rate of about 96%.

Fig. 7(a) Detection on a crowded image, (b) threshold of 0.7, (c) threshold 0.8, (d) Plot of S_e score given by ellipse area criterion with respect to azimuth angle.

7. CONCLUSION

In this paper, we present a novel method for pose invariant face detection. The novelty in our method of face detection arises from the integration of evidence from various sources such as shape, color and frequency response. By integrating these various evidence sources, a final score is obtained. Results of faces from frontal to profile views is shown and the detection rate is about 96%.

REFERENCES

[1] Computer Vision Laboratory, Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia and SCV, PTERS, Velenje, http://www.lrv.fri.uni-lj.si/facedb.html
[2] Christophe Garcia and Georgios Tziritas, "Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis," IEEE Transactions on Multimedia Vol. 1, No. 3, pp. 264-277, September 1999.
[3] Michael J. Jones, James M. Rehg, "Statistical Color Models with Application to skin Detection," Computer Vision and Pattern Recognition(CVPR 99), Ft. Collins, CO, pp. 274-280, June, 1999.
[4] Haiyuan Wu, Qian Chen and Masahiko Yachida, "Face Detection from Color Images using a fuzzy pattern matching methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 6, pp. 557-563, June 1999.
[5] Martin Riedmiller, Heinrich Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm," In proceedings of IEEE Intl. Conference on Neural Networks, 1993.59,93.
[6] Henry A. Rowley, Shumeeth Baluja and Takeo Kanade, "Neural Network Based Face Detection," IEEE Pattern Analysis and Machine Intelligence, Vol. 20, pp. 22-38, 1998.
[7] Fu Jie Huang, Zhihua Zhou, Hong-Jiang Zhang, Tsuhan Chen, "Pose Invariant Face Recognition," IEEE Intl. Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000, pp. 245-250.
[8] R. Brunelli, T. Poggio, "Face Recognition through Geometrical Features," European Conference on Computer Vision, pp. 791-800, 1992.
[9] H. Wang and S. F. Chang, "An highly efficient system for automatic face region detection in MPEG videos", IEEE Transactions on Circuit Systems for Video Technology, Vol. 7, No. 4, pp. 615 - 628, 1997