Proceedings of the 6th IASTED International Conference
COMPUTER GRAPHICS AND IMAGING
August 13-15, 2003, Honolulu, Hawaii, USA

HEURISTIC-BASED AUTOMATIC FACE DETECTION

Geovany Ramírez[1], Vittorio Zanella[1,2], Olac Fuentes[2]

[1]Universidad Popular Autónoma del Estado de Puebla
21 sur #1103 Col. Santiago Puebla 72160, México
[2]Instituto Nacional de Astrofísica Optica y Electrónica
Luis Enrique Erro #1 Sta. María Tonantzintla Puebla 72840, México
E-mail: in00020@upaep.mx, vzanella@upaep.mx, fuentes@inaoep.mx

ABSTRACT

Face detection is the first step for very important applications such as: face recognition, computer-human interface, and surveillance. This task is very difficult because each face is different and many factors can vary, including expression, pose, and light conditions. In this paper, we present a simple method based on heuristics to detect faces in grayscale images with complex backgrounds. The method is divided in two stages, the first stage searches for possible faces in the image, and the second stage determines if a possible face is really a face using two discriminators.

KEY WORDS

Imaging and Image Processing, Computer Vision, Automatic Face Detection.

1. INTRODUCTION

Face detection is the first step for applications as: face recognition, computer-human interface, and surveillance. The performance of these applications is limited by their capacity to detect faces quickly and accurately. Face detection is a difficult task because the face of each person is different, moreover, the face can vary in size, orientation and pose; it is also affected by the lighting conditions as well as the capacity of the device used to obtain the image.

Some approaches perform face detection in two steps. The first step can be used to determine the candidates to be a face using a fast algorithm [1], or to apply some filters to improve the face conditions, or to perform a pre-segmentation of the image [2, 3]. In the second step, more robust but more complex methods are applied to determine if the possible face is really a face and to refine the location, or to apply a complex algorithm in a specific regions. Using two steps, one simple and one complex, it is expected that all resources are applied only where necessary. In some occasions it is necessary to apply more than two steps, so is the case of Viola and Jones [4] that uses up to 32 steps, called classifiers; the first is very simple and fast but with a low accuracy and the last is very complex but with high accuracy.

The system of Heisele, Serre, Pontil and Poggio [5] based on face components, uses a two level hierarchy classifier to detect faces. In the first level the system only detects the components of the face (eyes, nose, mouth, etc.) in a grayscale image. In the second level another classifier uses the components detected in the first level to match with a geometrical model of the face. We based part of our work in two components of the face: the eyes and the nose, measuring the relationship between them.

In this paper, we present an approach to solve the problem of face detection using heuristics and a small set of faces. Our method works with images in grayscale and detects faces with a frontal pose and with uniform lighting. The method uses two steps called stages. The first stage looks for possible faces in the image using a simple heuristic*: In a face, the average intensity of the eyes is lower than the intensity of the part of the nose that is between the eyes.* This heuristic is a feature that Viola and Jones [4] had already obtained using AdaBoost with a training dataset of 4916 faces. In the second stage, it takes the possible faces and determines if a possible face is a face. In this stage two discriminators are used, the first one is based in a simple heuristic too: *The histograms of the image in grayscale of a face with a uniform lighting have a specific shape*. The second is based on the edge detection of the possible face.

The main contribution of this paper is a simple method to detect possible faces that works with images that have complex backgrounds.

2. THE SET OF FACES

Rowley, Baluja and Kanade [2] used 1,050 faces and 8,000 non faces to train their neural network. In the case of Garcia and Delakis [3], it used 12,976 faces and 15,000 non faces to train their neural network. Viola and Jones [4] used 4916 faces to obtain the features of a face. These sets of faces are very big in comparison with our set that consists of only 40 faces. We don't use our set to train our system, rather, it is used to check our heuristics and determine the acceptation thresholds.

The faces in our set are under the following conditions:

- *The face is in a frontal pose.*

- *The face doesn't have any extra structural components such as beards, mustaches, or glasses that can occlude the basic elements (eyes, mouth, nose, cheeks and eyebrows).*
- *The lighting is uniform on the face.*

Each image in the set contains only the face region; the size varies from 50x50 up to 200x200. All the faces are under ideal conditions, but the brightness varies from low intensity to high intensity. The images were obtained from: scanned photographs, digital cameras, the web and different test datasets including BioID [1], CMU+MIT [2] and ORL [6]. (See Fig. 1).

Fig. 1. Some examples from our set.

3. THE FIRST STAGE

The eyes are a very important part of the face, they show every emotion, but don't change drastically as in the case of the mouth. We use this principle to focus all the attention in the eyes because they have low variability. It would be difficult to perform a search of eyes in an image of a face in high resolution; this is rather an object recognition task that requires many resources. For this reason, it is convenient to work with face images in low resolution, because it only preserves the most relevant information.

3.1 The Heuristic

We propose the following heuristic:

In a face with uniform lighting, the average intensity of the eyes is lower than the intensity of the part of the nose that is between the eyes.

We will refer to the heuristic as *Eyes-Nose differential (ENdif)*. The heuristic above was formulated from the observation of faces from different distances. This was possible with a simple experiment. We reduced the size of our set of faces to 30x30, 20x20, 10x10 and 9x9 using common methods for resizing as 'nearest', 'bilinear', and 'bicubic'; after that histogram equalization was applied to make a simple normalization of the intensity of the pixels and to enhance the differences (See Fig. 2).

(a) (b) (c) (d) (e) (f)

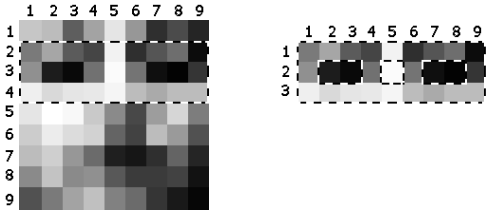
Fig. 2. (a) and (d) Original size. (b) and (e) Image resized to 9x9 size using Bilinear method. (c) and (f) After applying histogram equalization.

In 9x9 size, it is difficult to determine many of the features in the face, but the size is enough to locate the eyes. In this size, it is easy to see that the eyes are about 2x1 pixels and are darker than the part of the nose between of the eyes which is the central pixel. This heuristic is a feature that Viola-Jones [4] had already used before.

3.2 Eyes-Nose differential.

To apply the heuristic to an image of a possible face first it is necessary to resize the image to 9x9 pixels. The region of the eyes in a 9x9 image corresponds to rows 2, 3 and 4. Then the eyes region has a size of 9x3 (See Fig. 3). The equation to compute the *ENdif* from the eyes region, which we will call *I* is:

$$ENdif = I_{5,2} - \left(I_{2,2} + I_{3,2} + I_{7,2} + I_{8,2}\right)/4 \qquad (1)$$



(a) (b)

Fig. 3. (a) Face region of size 9x9 pixels; and the eyes region (framed with dashed line) corresponds to rows 2, 3 and 4. (b) Eyes Region *I* taken from (a).

3.3 Two Strategies

It is necessary to perform the search on the whole image to locate all the possible regions that can contain a face. The size of the regions depends on the size of the face to look for. It is necessary to determine the value of the *ENdif* of each region using (1). The regions that have an *ENdif* > *thENdif* (Eye-Nose differential acceptation threshold, calculated with our set of faces) are the regions that will be selected for the second Stage. The above mentioned can be performed in 2 ways, which are described in Sections 3.3.1 and 3.3.2.

We use the distance between the centers of the eyes (*ED*) as a parameter of the size of a face. To detect faces with different sizes, the value of *ED* varies, being increased by a factor of 1.15. The value of ED is used to calculate the size of the eyes region, as well as to create a pyramid.

3.3.1 First Method

The first method consists of using a search window of size *X_eye*×*Y_eye* (See Fig. 4), corresponding to the eyes region, to examine the original image. The values of *X_eye* and *Y_eye* vary depending of the size of the face to look for. Each region framed by the search window is clipped and resized to 9x3; then histogram equalization is

applied to determine its *ENdif* value. The values of *X_eye* and *Y_eye* are determined for the value of *ED*.

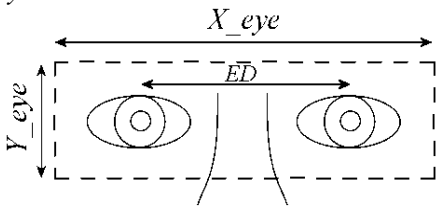


Fig. 4. The eyes and the eyes distance (*ED*). The eyes region is determined using equations (2) and (3).

To determine the values of *X_eye* and *Y_eye*, we take the *ED* value and the aspect ratio of the eyes region that is 3.3 (this aspect ratio is used because it offers better results than using an aspect ratio of 3). In a face of size 9x9 pixels the value of *ED* is 5, and the value of *X_eye* is 9 (see. Fig. 3). With these data we can get the following relationships:

$$X\_eye = 1.8ED \quad (2)$$

$$Y\_eye = X\_eye / 3.3 \quad (3)$$

It is not necessary to perform pixel-by-pixel search, unless the face, smaller than 30x30. Instead, the window can jump an *inc* amount of pixels horizontally, as well as vertically. This increases the speed of the search. The value of *inc* is calculated with:

$$inc = Y\_eye / 3 \quad (4)$$

All the regions that have an *ENdif* greater than *ThENdif*, will save their locations for the second stage.

3.3.2 Second Method

The second way works with a pyramid of images. This pyramid is created using *ED* as a parameter to determine what percentage of the original image should be resized to transform a face with any *ED* to a face with an *ED*=5. All the images in the pyramid are resized from the original image using a bilinear method.

A fixed search window examines each image in the pyramid clipping a region of 9x3 and applying histogram equalization to determine its *ENdif*. All the regions that have an *ENdif* greater than *ThENdif* will save their locations for the second stage.

3.4 Merging the Two Strategies

Experimental results show that the first method obtains the exact position of the possible faces, however it is a slow process; on the other hand the second method obtains an approximate position quickly. Moreover the first method can eliminate more irrelevant information than the second, because it uses higher acceptation thresholds and it is more accurate. Based on the above we merged the two methods, using the first one to look for all the possible faces with an approximate position and the second to refine the position and to begin eliminating false faces.

4. THE SECOND STAGE

The regions selected by the first stage correspond to the eyes region of the possible faces; these regions are extended to the size of the face region. To determine if the possible face is really a face we apply the first discriminator, which is based on a heuristic. If the possible face is accepted by the first discriminator, then the second discriminator is applied.

We will define a discriminator as a function that takes an image of a possible face and returns 2 values; one is a binary value: "yes" or "no", answering the question: Is it a face?, the other is the adjustment value. The binary value is calculated comparing the adjustment value against *th_dis* (discriminator acceptation threshold); each discriminator has a different *th_dis* calculated with our set of faces.

4.1 The Heuristic

We know that a face has eyes, nose, mouth, eyebrows and it is covered by skin, and some elements of the face are darker than others. The relationship between the elements of the face results in a histogram with a specific shape. We propose the following heuristic:

The histograms of the image in grayscale of a face with uniform lighting always have a specific shape.

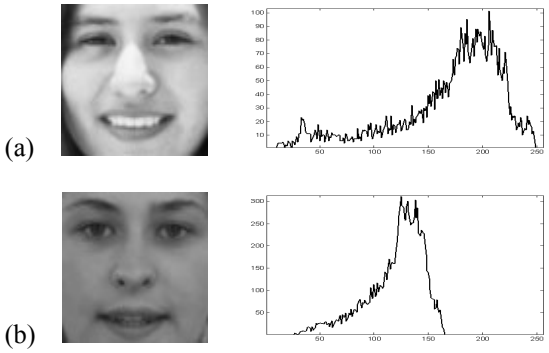


Fig. 5. Example of the histograms from two face images. The shapes of histograms (a) and (b) are similar, one is wider than the other one and the values are different, this due to the lighting conditions and the different sizes in the images.

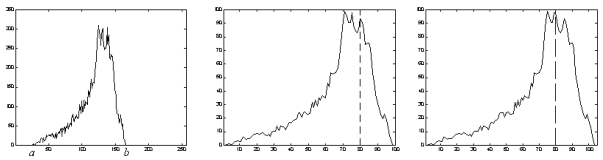
The heuristic above was tested with the histograms from our face set (See Fig. 5). Although the images contain faces under ideal conditions, this heuristic is applicable practically in any face image under the conditions described below:

- *All the components of the face should be visible (eyes, nose, mouth) regardless of whether the face is rotated.*
- *The image should not have any previous processing that affects the histogram drastically e.g. Histogram equalization.*
- *Lighting should be uniform on the face.*
- *The image should be acquired with enough color (e.g. not have an appearance of dithering or posterized).*

4.2 Discriminator 1

The discriminator uses a curve model to compare it with the histogram of the possible face. Due to the variability of the size and lighting conditions of the possible faces, their histograms can vary in magnitude and in number of elements. So it is necessary to normalize the histogram of the possible face.

The normalization scales the magnitude of the histogram to values between 0 and 99. A new histogram is created taking only the part of the histogram corresponding to the range [*a*, *b*], where *a* is the position of the first element with a value higher than 0, and *b* is the position of the last element with a value higher than 0. Then the new histogram is resized to 100 elements. It is also necessary to move the peak of the new histogram to position 80, in this case the part of the new histogram corresponds to the elements in position 0 up to the average of all the peak, then it is resized to 80 elements and the rest is left with the same size (See Fig. 6).



(a) (b) (c)

Fig. 6. (a) Original histogram of the face image in Fig. 5b. (b) The new histogram normalized in magnitude and elements. (c) The new histogram with the peaks average in position 80.

Using the software "TableCurve 2D" for curve fitting, it was determined that the shape of the average of the normalized histograms of our set of faces corresponds to a Pearson type IV distribution. We use this distribution to build the model curve of the shape (See Fig. 7).

The discriminator compares the normalized histogram of the possible face with the curve model using the coefficient of determination (r^2). The coefficient of determination is the percentage of the variation of the normalized histogram that can be explained by the model, compared with the mean value of the normalized histogram. The value of r^2 is the value of adjustment that discriminator 1 returns, if the value of r^2 greater than *th_dis_1*, then the possible face is evaluated by discriminator 2.

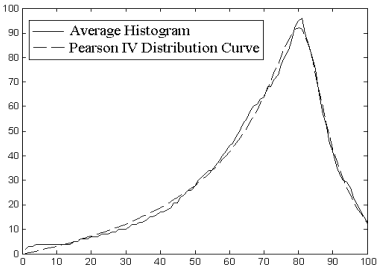


Fig. 7. Pearson IV distribution curve and the average normalized histogram from our set of faces.

4.3 Discriminator 2

This discriminator first segments the image using the Sobel method for edge-finding in horizontal direction; after that, dilation is performed to enhance the edges. The resulting image will contain only the eyes, the eyebrows, the mouth and the nose; we use a simple mask to evaluate the presence or absence of these elements (See Fig. 8). The image is resized to 30x30 pixels, which is the size of the mask. The mask is divided in 7 regions corresponding to the elements of the face, 1 and 2 correspond to the eyes, 3 and 6 correspond to the nose, 4 and 5 correspond to the cheeks, and 7 correspond to the mouth. A representative value is calculated for each region obtaining their average. In the case of 1, 2, 6 and 7, the expected representative value should be near 0, and in the case of 3, 4, and 5, the expected value should be near 255. Symmetry is also evaluated between the regions 1 and 2, and the regions 4 and 5, comparing their representative values. For each possible face 7 conditions are evaluated:

- The average value in regions 1 and 2.
- The difference between regions 1 and 2.
- The average value in regions 4 and 5.
- The difference between region 4 and 5.
- The values of region 3.
- The values of region 6.
- The values of region 7.

(a) (b) (c)

Fig. 8. (a) The mask used by Discriminator 2. (b) After performing edge detection, applying a dilation effect and resizing the image of the possible face. (c) Using the mask to evaluate if the possible face is really a face.

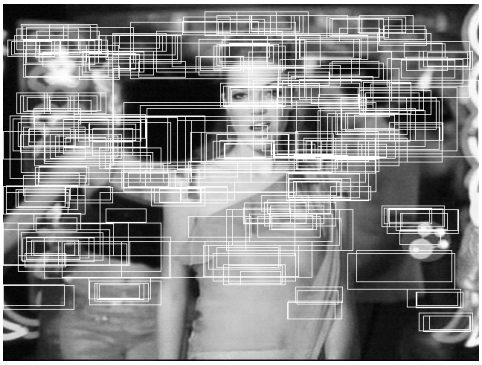
We used our set of faces to determine the representative values of each region that is expected. The adjustment values are chosen so that if all the conditions are fulfilled, the value is 24. If the adjustment value is

over *th_dis_2* (obtained with our set of faces) then the possible face is accepted as a face.

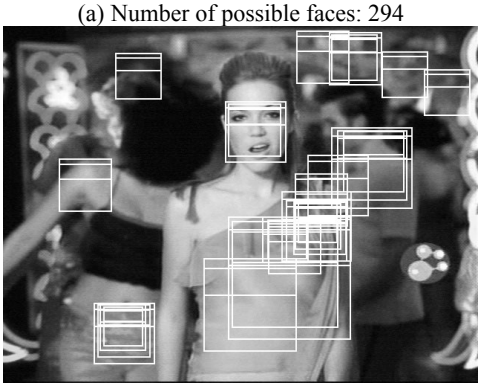
When the whole process is finished, often a face is detected in different scales but only one scale is required per face. In this case the overlapping detections are clustered and the final bounding corresponds to the detection with best values. This is applied to each face.

5. EXPERIMENTAL RESULTS

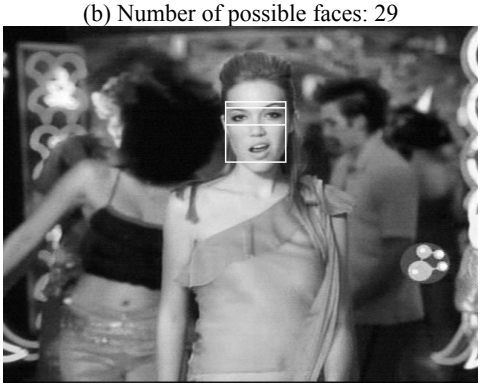
We experimented with different images and with different faces sizes. One example of the whole process is presented below (Fig. 9) with an image of 640x480 pixels.



(a) Number of possible faces: 294



(b) Number of possible faces: 29



(c) Number of faces detected: 1

Fig.9 The process to detect a face. (a) The first stage detected 294 faces of 24646 windows. (b) After discriminator 1. (c) After discriminator 2 and clustering.

5.1 Acceptation thresholds values

The acceptation threshold values were statistically determined using our system with our face dataset. For the *thENdif* (Eye-Nose differential acceptation threshold) we used two values, the first one for the approximate detection is 115, and second for the refined detection is 150. The value of *th_Dis_1* (Discriminator acceptation threshold) is determined depending of the level of accuracy that is required and the set of faces used, a value greater than 65% will give good results. The value of *th_Dis_2* is 10.

5.2 BioID Face Database

Our method was tested with the test set from BioID Face Database [1]; this face dataset consists of 1521 grayscale images of size 384x288 pixels, with a frontal view of 23 different persons' faces. This set has a very large variety of illumination, background and face sizes. Since many of the images in this dataset are not in ideal conditions (very dark, non uniform lighting and glasses that occlude the eyes) we use only the first 1000 images that have mainly good conditions. The result is 87.3% of correct detections with 41 false detections. The faces were not detected because of the following reasons: the image didn't have a uniform illumination (which affects the histogram), the image was extremely dark (which also affects the histogram), or the faces had glasses that reflected light (which affects the histogram and the edges).

6. CONCLUSIONS AND FUTURE WORK

Our system can work very well with images of faces under ideal conditions but it also has a good performance with faces in condition near ideal. The first stage can eliminate around 99% of all the possible regions of an image and yields a correct detection of around 98% of the faces in BioID Face Database. It can detect faces rotated up to 5 degrees. This stage is limited for frontal faces, but it is possible to determine another search window based on the same heuristic but oriented to semi profile faces.

The second stage has a good performance with faces with enough and uniform lighting. The first discriminator practically works with any image under the conditions described in Section 4.1, which are out of the ideal conditions. It also eliminates around 90% of the possible faces detected by the first stage.

Each stage can be used separately for diverse applications; maybe the first stage can be used for pre-segmentation of the images, to apply a very robust method such as [2] or [3] afterwards.

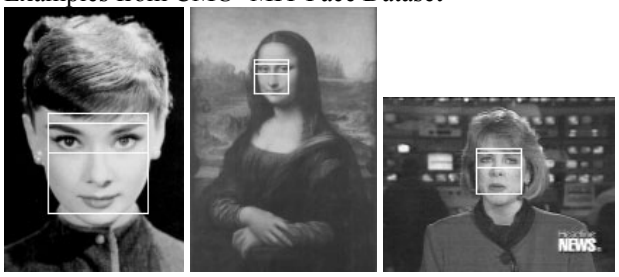
This was a first approach to solve the problem of face detection using simple heuristics; the system works very well under ideal conditions, future work will be oriented to improve the capacity of the two stages. In the first stage, it will be to improve its capacity to detect faces in a

semi profile pose and rotated by any amount of degrees. For the second stage, in the second discriminator, it will be to look for a better way to compare the edges of the face.

There are many applications of face detection, but this system can be used to perform the first step in applications with a controlled environment (mainly in the lighting condition), such as surveillance or in a biometric authentication.

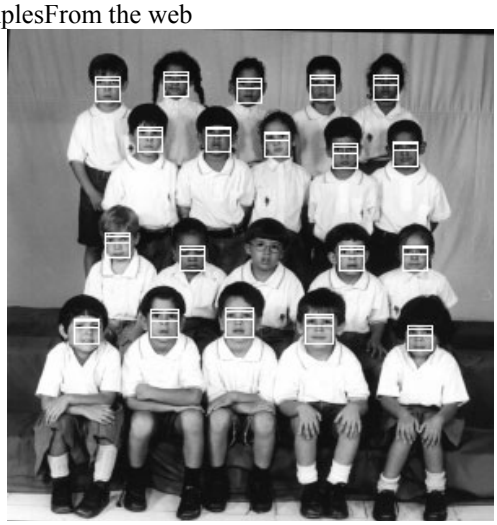
Examples from BioID Face Dataset

Examples from CMU+MIT Face Dataset



Examples from our set of faces (left) and from our test set (right).

ExamplesFrom the web



REFERENCES

[1] O. Jesorsky, K.J. Kirchberg, and R.W.Frischholz, Robust Face Detection Using the Hausdorff Distance, *In Proc. Third International Conference on Audio- and Video-based Biometric Person Authentication, Springer, Lecture Notes in Computer Science,* LNCS-2091, pp. 90–95, Halmstad, Sweden, 6–8 June 2001.

[2] H. Rowley, S. Baluja, and T. Kanade, Neural network-based face detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *volume 20*, no. 1, pp. 23-38, January 1998.

[3] C. Garcia and M. Delakis, A neural Architecture for Fast and Robust Face Detection, *Pro. IEEE-IAPR International Conference on Pattern Recognition* (ICPR'02), Quebec city, Canada, August 2002.

[4] P. Viola and M. Jones, Robust real-time object detection, *Second International Workshop on Statistical and Computation Theories of Vision-Modeling, Computing, and Sampling,* Vancouver Canada, July 2001.

[5] B. Heisele, T. Serre, M. Pontil and T. Poggio, Component-based Face Detection, *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 2001), Kauai, Hawaii, Vol. 1, 657-662, December 2001.

[6] F. Samaria and A. Harter, Parameterisation of a stochastic model for human face identification. *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota (Florida), December 1994.