FACE DETECTION UNDER ARBITRARY LIGHTING CONDITIONS AND BACKGROUND ENVIRONMENT

DIONYSIS ANIFANTIS, EYANGELOS DERMATAS, GEORGE KOKKINAKIS
Department of Electrical & Computer Engineering
University of Patras
Patras, Kato Kastritsi, 26500
HELLAS

Abstract: - This paper presents and evaluates a neural network method for detecting human faces in arbitrary gray scale images. Some novel techniques are used to face the problems of the illumination conditions, the facial peculiarities and rotation in arbitrary images.

The illumination distortion of the face features is minimized by splitting each candidate rectangular window into two vertical, equal-size, successive and non-overlapping sub-images and by applying non-linear equalization to each window independently. A multilayer percepton neural network is used to detect faces under various facial expressions, skin color, presence of glasses, mustache, beard etc. This performance is achieved by training the network weights using an adequate set of various faces under a wide range of face lineaments. Negative examples are included in the training set by adding false detected faces in a set of non-face images during the training process. The face orientation is detected by rotating the image in a specified range.

The experiments were carried out in the Olivetti, Yale and Ackerman's databases. The proposed face detection method gave 96.91% detection accuracy in a set of 1199 frontal view faces.

Key-Words: - Face detection, light correction, neural networks, neural face detector, face orientation, boot-strap method.

1 Introduction

The importance of face processing in visual scenes has steadily increased over the last decade. Computer detection and analysis of the human face can be used in a great number of human-machine communication applications such as visual speech recognition, identification of an authorized person, recognition of facial expressions, etc [1,5]. Faces are similarly structured but different perspective variations, rotation positions, background environment and various lighting conditions distort the image, producing a wide distribution of the gray scale, and the statistical texture-based features [4,8]. Template matching and geometrical object recognition methods tend to fail by detecting faces in arbitrary lighting [8,9]. A change in light source distribution can cast or remove significant shadows from a particular face, and in different background conditions [10].

Advance training methods have been used to incorporate the wide distribution of the frontal view of face patterns in neural network knowledge. Recently, Rowley at al. [3,6] have shown that neural networks can handle both facial peculiarities and illumination conditions. The face database can be artificially enlarged so that a more comprehensive sample is obtained, by rotating each face example in the range of 10 degrees [3]. In this case the neural network has the capability to detect images under various orientation positions but the orientation information has been incorporated into the network knowledge, therefore the rotation angle can not be reached in this approach.

An efficient solution to the selection problem of the negative examples (non-face images) has already been presented in the direction of minimizing the false acceptance error [6]. This technique reduces the number of falsely recognized faces in arbitrary images, a critical factor for developing practical face recognizers.

In this paper we present and evaluate a novel neural network method for accurate detection of frontal view faces in arbitrary gray scale images.

Initially, multiple images are created from the candidate image by rotation in a step of approximately 3 degrees and scaling by a factor of 0.8. The FRs are detected by scanning each image using window analysis in an overlapping basis of 2-pixels. This technique is computationally costly but it permits FR detection at any size, position and orientation.

In each image a two-level process is applied; the window normalization and the neural network classification. Several image window hypotheses are generated and characterized as face regions (FRs).

In the window normalization process each sub-image is split into two vertical, equal-size, successive and non-overlapping sub-images. Each sub-image is normalized using advance illumination filters. In this first-level process, the proposed light normalization method minimizes the features dispersion of the illumination phenomenon even in cases where a strong light source is positioned next to a human face.

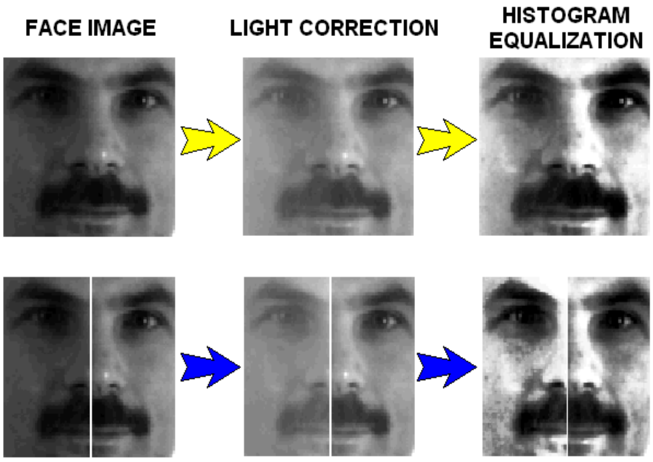

Fig. 1 Example of light and histogram equalization

In Figure 1 non-linear illumination distortion appears in a face image due to the presence of a light source on the right side of the camera. Two different light and histogram equalization methods are applied to the same image. In the first case light correction and histogram equalization are successively applied to the original image, as shown in the first three images of Figure 1. In the second case the first image is divided into two sub-images and after that the same transformations are applied to each independent sub-image, as shown in the last three images of Figure 1.

A neural network classifier detects the presence of a face for each normalized window by mapping the input vector space into the two-dimensional space of the output neurons. Each output value represents the belief level of the neural network on the presence or absence of a face. The proposed structure of the multilayer perceptron neural network increases the computational complexity but it faces the problem of recognizing different face features (skin color, presence of a moustache, beard, glasses, etc).

Extensive experiments have shown that the FR accuracy of the neural detector is improved significantly in case that the training data are preprocessing by using non-linear equalizers of illumination. Moreover, additional improvement is achieved by performing orientation calibration of the training faces. The proposed method has been evaluated in a test-bed of 1199 faces (a unification of the Olivetti, Yale, a subset of the FERET database, and a collection of face images from Bern University). The experiments have given a mean face detection rate of 96.91%.

The rest of this paper is organized as follows. In the next section the proposed method is described in detail. Section 3 presents the face image database used for the evaluation of the method. In section 4 the experimental results are given and finally a discussion on the current directions of our research concludes this work.

2 Face detection method

The proposed face detection method can be separated into the window normalization and the neural network classification process, as shown in the flowchart of Figure 2.

2.1 The window normalization process

The aim of this process is to eliminate the illumination variability for all possible face positions and to adjust the dimensionality of the processed data to the neural network input vector. The following image transformations are performed:

- Rotation of the original image creates multiple images.
- The proposed method's capability of detecting faces of any size is achieved by reducing each image repeatedly. In our experiments the image is downscaled by a factor of 1.2.
- The capability of detecting faces anywhere in the image is achieved by scanning the whole image with a 30x30-pixel window in an overlapping mode having a step of 2-pixels.

Each 30x30-pixel image is transformed by performing masking, non-linear illumination and histogram equalization.

2.1.1 Masking

A 30x30-pixel mask is used to eliminate some near-boundary pixels of each window pattern. Eliminating these pixels ensures that we don't wrongly encode any unwanted background pixels. This masking process adjusts the image data to the dimension of the neural classifier input vector.

First we construct the oval mask (Figure 3) and then for each image we multiply the mask and the window data to produce the final masked image.

So the values that are inside the oval region remain untouched while the rest values are eliminated. The total number of processed pixels decreases from 900 to 780.

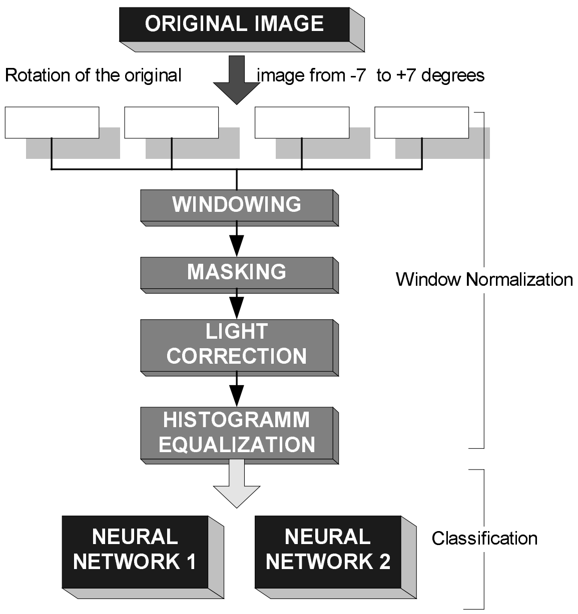

Fig. 2 Flowchart of the Face detection method

2.1.2 Light Correction

The amount of light scattered to the viewer or camera from each section is a function of the orientation of the surface with respect to the source of light and the viewer, even if the surface material and finish are uniform. Fortunately, in most of such situations the variation of background brightness is a smooth and well-behaving function of location and can be approximated by simple functions such as polynomials.

The proposed light correction method reduces heavy shadows caused by extreme lighting conditions. In Figure 1 we show that it is better to

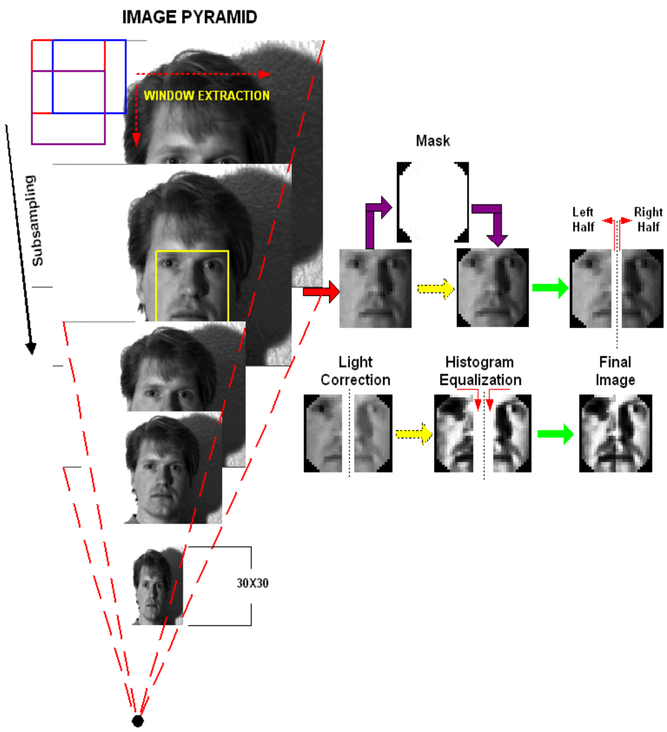

Fig. 3 Feature Extraction Process

split each window into two vertical, equal-size, successive and non-overlapping sub-images and subtract a best-fit brightness plane from each section, rather than subtracting a brightness plane for the whole-unmasked window.

We first divide the 30x30-pixel image into two regions of 15x30 pixels corresponding to its left and right half section (Figure 3). By selecting the pixels inside of the oval mask, for each of the two sections, a list of brightness values and locations can be acquired. These pixels are used to perform least-squares fitting to the illumination background function. The constructed best-fit brightness plane for each section (left - right) is then subtracted from the unmasked window pixels.

2.1.3 Histogram Equalization

This normalization operation expands the range of intensities in each section inside the oval mask. This compensates for differences in camera response curves and improves contrast in some cases. In Figure 1 we show the result of performing histogram equalization separately in each of the two sections.

Histogram equalization is a non-linear process reassigning the brightness values of pixels based on the image histogram. Individual pixels retain their brightness order (each pixel remain brighter or darker than other pixels) but the values are shifted, so that an equal number of pixels have each possible brightness value. In many cases, this spreads out the values in areas where different regions meet, showing details in areas with a high brightness gradient. The process is quite simple and consists of four steps.

1. The running sum of the histogram values is estimated.
2. The sum acquired from the previous step is normalized with the total number of pixels.
3. The normalized sum is multiplied by the maximum gray level value and rounded.
4. The original gray level values are mapped to the results from Step 3 using one-to-one correspondence

2.2 The Neural network classifier

In the second process, face presence in each window is detected by using a two-layer perceptron classifier with 390 input gray-scale values, 16 nodes in the hidden layer and 2 output nodes. The Neural Network (NN) classifier is trained to recognize only the left half of a face. In order to detect the whole face the following technique is used:

1. We construct two NN's with the above structure, corresponding to the left and right half section of a face.
2. The NN for the right section of the image is constructed by mirroring the weights of the hidden layer of the neural network of the left section taking into account the fact that the frontal view of a face is symmetric (in the majority of cases).

The unmasked pixels of each section of the window are processed through the corresponding NN's activating four output nodes.

2.2.1 Face Detection algorithm

Faces are detected for all candidate windows satisfying one of the following conditions:

1. Both NN-outputs assigned to the face presence have greater activation values than the outputs corresponding to the absence of a face (case AND).

In case that this condition is not satisfied, the second condition is examined:

2. The sum of the activation values of the face presence outputs is greater than the sum of the activation values of the face absence outputs (case SUM).

These face detection conditions have been reached experimentally and seem to work satisfactorily, as is shown from the experimental results presented in this paper.

2.2.2 Training the face detection network

For training the neural network to serve as an accurate filter a large number of faces and non-faces, i.e. nearly 2578 face examples were used. The eyes, tip of nose, corners and center of the mouth of each face were labeled manually. These points are used to locate the exact face position and subsequently to normalize each face to the same scale, position and orientation. The alignment algorithm is adapted from [3].

The labeled faces were carefully collected having uniform background and almost all other facial features that the face detector will encounter afterwards.

Next, from the normalized face examples we selected the most illuminated section and applied the preprocessing filters. Mirroring transformation was performed in case where the right section of the face image was illuminated more than the left section. Thus, in all cases the training examples were left half section of the face images as shown in Figure 4.

The collection of non-face examples was achieved with the aid of a bootstrap method [2] (Figure 5):

1. Construct an initial set of images, which contains no faces, by selecting randomly sub-images from a set of texture and scenery images.
2. The neural network is trained to produce the desirable output (0.9,0.1) for face images, (0.1,0.9) for non-face images. Initially, the network's weights are set randomly in near to zero values. After the first iteration, the weights computed in the previous iteration are used as a starting point for the next iteration of the algorithm.
3. Using the current neural weight configuration, the incorrectly detected face windows in the images that contain no faces are collected.
4. The collected windows are added into the training set as negative examples. Go to step 2 and repeat the procedure until the classifier eliminates the incorrectly detected faces.

In our experiments 300 texture images and 200 scenery images were used for collecting negative examples. The presence of these examples forces the classifier to learn the precise boundaries between face and non-face images. A typical training run selects about 11000 non-face images.

3 Face databases

A large set of images was processed in the experiments. These images were collected from WWW-databases, so that they featured differences in lighting conditions, facial expressions, and complex backgrounds. In the following we present briefly the testing material. All images of the following face databases contain only one face per image.

1. *The ORL database from the Olivetti Research Laboratory in Cambridge, UK*. There are 10 different images of 40 different persons and the size of each image is 92x112-pixels, giving a total number of 400 images. Variations in facial expressions (open/close eyes, smiling/no-smiling), facial details (with glasses), scale (up to 10%), and orientation (up to 20 degrees) are met in this database.

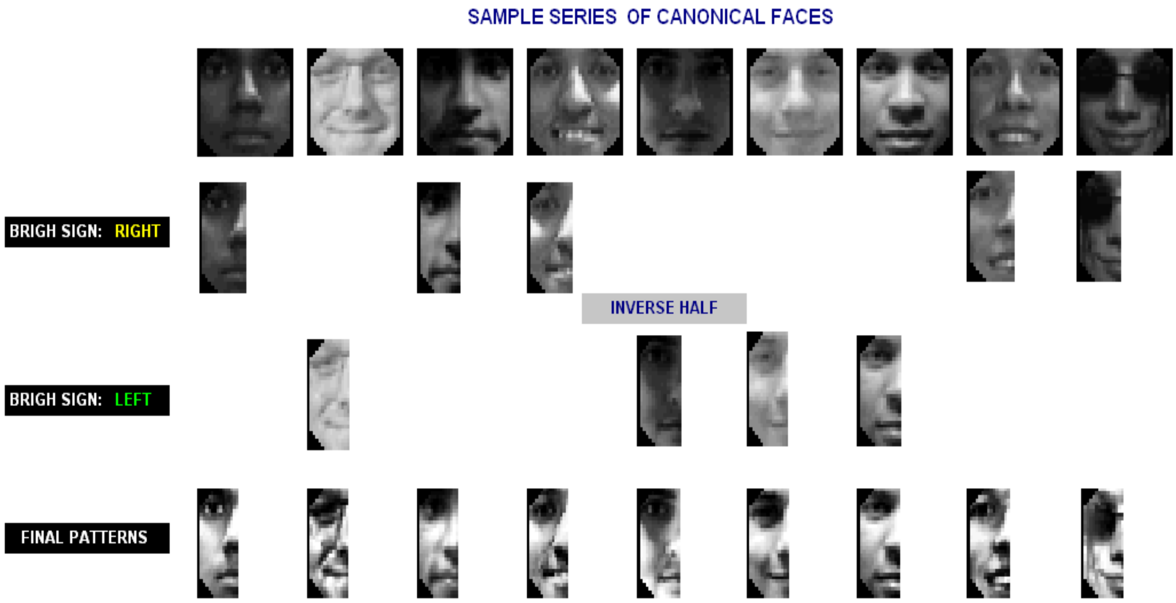

Fig. 4 Building left section examples of Faces.

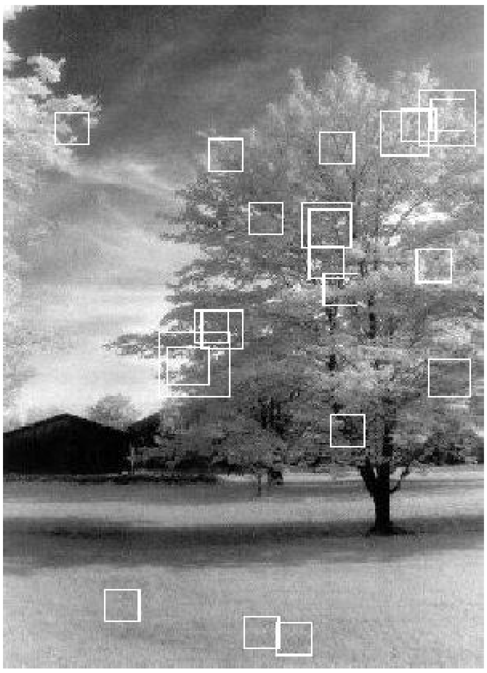

Fig. 5 Collecting Negatives Examples from scenery image.

2. *The Yale's database*. Contains 165 images (320x243-pixels) of 15 subjects. There are 11 images per subject, one image for each of the following facial expressions or configurations: center-light, happy, left-light, yes/no glasses, normal, right-light, sad, sleepy, surprised, and wink.
3. *The Ackerman's database from the University of Bern, Switzerland [7]*. This database contains frontal views of 30 people. For each person there are 10-greylevel images (512x342-pixels) with slight variations of the head positions: right into the camera, looking to the right, looking to the left, downwards, upwards (300 face images).
4. A subset of the FERET database. Each image (128x192-pixels) contains one face with (in most cases) a uniform background and good lighting. The total number of the set images is 334.

4 Experimental Results

In Table 1 we present the experimental results which were achieved using only the first condition (case AND). The face detector has a small number of false positives, but it also ignores a great number of faces. The Yale's database gave the minimum detection accuracy due to its extreme light conditions and facial expressions.

Table 1. Detection accuracy using condition AND

RESULTS		AND		
Face Databases	Images	Miss Faces	False Positives	Detection Rate
Olivetti (ORL)	400	43	1	89.25%
Yale's Database	165	39	4	76.36%
Bern Database	300	15	7	95.00%
FERET (Subset)	334	15	0	95.51%
TOTAL	1199	112	12	90,7%

In Table 2, the face detection accuracy of the proposed method is given. The additional rule improves significantly the overall performance by eliminating the number of falsely rejected faces and increasing the falsely accepted window candidates. Applying simple heuristic rules can further eliminate the latter errors as Rowley et al. proposed [3].

Table 2. Accuracy of the face detection method

RESULTS	AND & SUM		
Face Databases	Miss Faces	False Positives	Detection Rate
Olivetti (ORL)	8	40	98.00%
Yale's Database	24	88	90.90%
Bern Database	3	12	99.00%
FERET (Subset)	2	8	99.40%
TOTAL	37	148	96,91%

5 Conclusion

The experimental results have given a mean face detection rate of 96.9% in a set of 1119 images for the proposed neural based face-detection system.

Currently our research is concentrated in a number of system restrictions. Specifically, the proposed method detects faces looking onto the camera. Geometrical models can be used to identify faces at different head orientations. An obstacle to incorporate the proposed method in practical systems is its computational complexity. A simpler face detection method can be used to detect regions of suspicion (ROS) and then the proposed method can be used within ROS regions for fine-tuning. Color segmentation methods can be used also for the detection of ROS, reducing significantly the false acceptance regions.

References:

[1] Obrecht R., and Jacob B. *Direct identification vs. correlated models to process acoustic and articulatory informations in automatic speech recognition*. ICASSP-97 pp. 999-1002

[2] Kah-Kay Sung. *Learning and Example Selection for Object and Pattern Detection.* Ph.D. thesis, MIT AI Lab, January 1996

[3] Rowley H., Baluja S., and Kanade T. *Neural network-based face detection*. Trans. on Pattern Analysis and Machine Intelligence, Vol.20, No.1, Jan. 1998, pp.23-38

[4] Rae R., and Ritter H. *Recognition of human head orientation based on artificial neural networks.* IEEE Tr. on Neural Networks Vol.9, No.2, Mar. 1998, pp. 257-265

[5] Wang C., and Brandstein M. *A hybrid real-time face tracking system.* ICAS SP-98

[6] Kah-Kay Sung and T. Poggio. *Example-based Learning for View-based Human Face Detection.* Trans. on Pattern Analysis and Machine Intelligence, Vol.20, No.1, Jan. 1998, pp.39-51

[7] From the World Wide Web (request: ackerman@iam.unibe.ch)

[8] McKenna S., Gong S., Wurtz R., Tanner J., and Banin D. *Tracking Facial Feature Points with Gabor Wavelets and Shape Models.* Proc. 1st Conf. On Audio-Video based Biometric Person Authentication, Lectures Notes in Computer Science, Springer Verlang, 1997

[9] Kin Choong Yow and Roberto Cipolla, *Feature-Based Human Face Detection,* Department of Enginnering University of Cambridge CB2 1P2 England.

[10] Dai Y., Nakano Y., and Miyao. *Extraction of facial images from complex background using SGLD matices*. Proc. Int. Conf. on Pattern Recognition, Jerusalem 1994, pp. 137-141