A NEURAL NETWORK METHOD FOR ACCURATE FACE DETECTION ON ARBITRARY IMAGES

D. Anifantis, E. Dermatas and G. Kokkinakis

Wire Communications Laboratory, Department of Electrical & Computer Engineering, University of Patras
Patras, Kato Kastritsi, 26500 HELLAS
Email: anifanti, dermatas@george.wcl2.ee.upatras.gr

ABSTRACT

In this paper we present a neural detector of frontal faces in gray scale images under arbitrary face size, orientation, facial expression, skin color, lighting conditions and background environment. In a two-level process, a window normalization module reduces the variability of the features and a neural classifier generates multiple face position hypotheses.

Extended experiments carried out in a test-bed of 6406 face images, have shown that the face detection accuracy is increased significantly when non-linear and probabilistic illumination equalizers pre-process the sub-images. Moreover, better results can be achieved in case of training the neural detector using positional and orientation normalized face examples. In this case the neural face detector has the capability to locate both position and orientation of a face.

In the multiple face position hypotheses generated by the proposed neural method, 98.3% detection accuracy, the highest reported in the literature, was measured.

1. INTRODUCTION

Human face detection methods have been studied for more than 20 years. The face identification problem in poor illumination conditions, different perspective variations, and background environment becomes a great challenge in real-life applications [1-3,9].

Faces are similarly structured with the same features arranged in roughly the same spatial configuration. Nevertheless, even among images of the same person's face, significant geometrical and textural differences are met due to changes in expression, illumination conditions and the presence of facial makeup. Therefore, traditional template matching and geometrical object recognition methods tend to fail by detecting faces in arbitrary lighting, a change in light source distribution can cast or remove significant shadows from a particular face, and in different background conditions. Neural detectors have been proposed recently and the experiments give extremely satisfactory results [4].

In particular, advanced training methods have been used to incorporate the wide distribution of the frontal view of face patterns in neural network knowledge [4,5]. An efficient solution to the selection problem of non-face images has already been presented aiming at minimizing the false acceptance error [4]. This technique reduces the number of falsely recognized faces in arbitrary images, a critical factor for developing practical face recognizers. The most accurate face detector has been reported recently giving recognition rates up to 96% by using a neural network method [4].

In this paper we present and evaluate an accurate neural detector of frontal view faces on arbitrary images. In a two-level process, the window normalization and the NN classification, a number of image window hypotheses is generated and characterized as face regions (FRs). In the window normalization process, each sub-image is transformed into a 30x30-pixel image and illumination filters minimize the dispersion of the illumination phenomenon. The neural network classifier detects the presence of a face by mapping the input vector space into the two-dimensional space of the neural network output. Each output value represents the belief level of the neural network knowledge 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 detecting the position and the orientation of faces with different features (skin color, presence of a moustache, bear, glasses, etc). The FRs are detected by scanning each image using window analysis in an overlapping mode.

The proposed method has been evaluated in a test-bed of 6406 faces (the Olivetti, and Yale databases, a subset of the FERET database, a collection of face images downloaded from the WWW and a set of 350 texture and 220 scenery images. The experiments have shown that the proposed FR detection method's accuracy is improved significantly in case of incorporating the non-linear equalizers of illumination and performing orientation calibration of the training faces, giving a face detection accuracy of 98.3%.

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 the last section the experimental results and the current directions of our research are given.

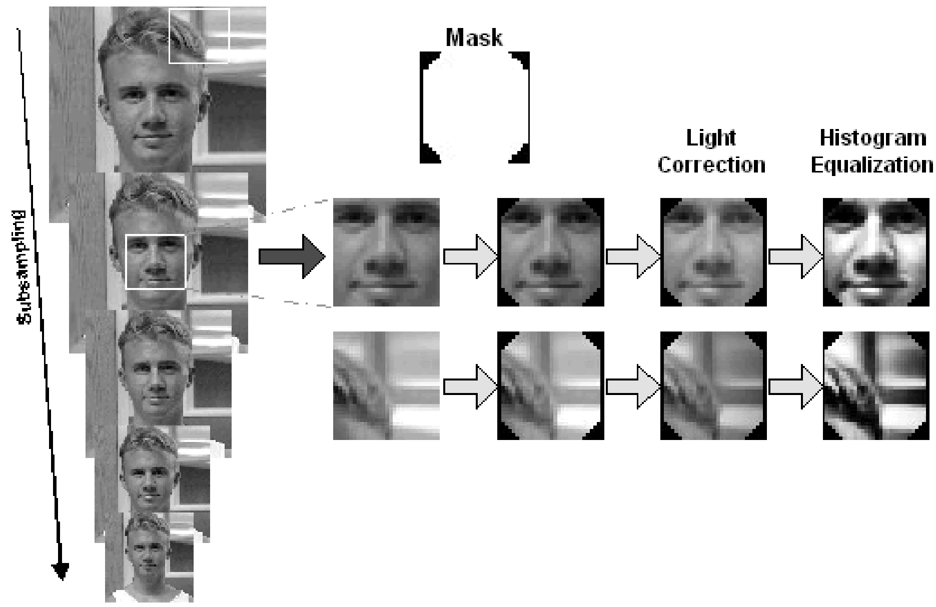

Fig 1: Filtering Process

2. FACE DETECTION METHOD

The proposed face detection method can be separated into two main processes, the window normalization and the classification.

2.1 The window normalization process

The target of this process is to eliminate the illumination variability for all possible face positions and to adjust the dimensionality of the processed data. 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 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.

By performing masking, non-linear illumination and histogram equalization each 30x30-pixel image is transformed, as follows:

Masking. We use the 30x30-pixel mask 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 (Fig. 1) 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. In this way, the total number of processed pixels decreases from 900 to 780.

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. We subtract a best-fit brightness plane from the masked window (oval region) pixels. For face patterns this does a fair job of reducing heavy shadows caused by extreme lighting conditions. By selecting the inside of the oval region pixels, a list of brightness values and locations can be acquired. These are then used to perform least-square fitting of the illumination background function. The constructed best-fit brightness plane is then subtracted from the unmasked window pixels.

Histogram Equalization. This normalization operation expands the range of brightness intensities in the masked window. This compensates for differences in camera response curves and improves contrast in some cases.

Histogram equalization is a non-linear process reassigning the brightness values of pixels on the basis of the image histogram. Individual pixels retain their brightness order (each pixel remains 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. This process is quite simple and consists of four steps: (a) the running sum of the histogram values is estimated, (b) the sum acquired from the previous step is normalized with the total number of pixels, (c) the normalized sum is multiplied by the maximum gray level value and rounded, and (d), the original gray level values are mapped to the results from Step (c) using one-to-one correspondence.

2.2 The Neural network classifier

A two-layer perceptron classifier (16 nodes in the hidden layer and 2 output nodes) detects face presence on each window by processing the 780 gray-scale unmasked pixels. A face is detected in the candidate window if the neural output assigned to the presence of a face gives a higher value than the output assigned to the non-face images.

2.3 Training the face detection network

For training the neural network to serve as an accurate filter we used a large number of faces and non-faces, i.e. nearly 2578 face examples. The eyes, tip of nose, corners and center of the mouth of each face were labeled manually. These points were used to normalize each face to the same scale, orientation, and position employing an alignment algorithm adapted from [4]. From each normalized face example we produced two more images by translating one pixel left – right, and reaching a total number of 7734 face patterns. The labeled faces were carefully collected having uniform background, and almost all-facial features that the face detector will encounter afterwards.

We collected also non-face examples for the training with the bootstrap method [6,8] in the following way:

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 and (0.1,0.9) for non-face images. Initially, the network's weights are set randomly in near to zero values.
3. Using the current neural weight configuration in the images without faces, the incorrectly detected face windows 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 350 texture images and 220 scenery images were used for collecting negative examples. The presence of these examples forced the classifier to learn the precise boundaries between face and nonface images. A typical training run selects about 11000 non-face images.

3. FACE DATABASES

Processing of a large set of images was carried out in the experiments. The image databases were collected from the WWW and featured differences in lighting conditions, facial expressions, and background. In the following we present briefly the testing material.

1. The ORL database from Olivetti Research Laboratory in Cambridge UK. There are 10 different images of 40 different persons and the size of each image is 92x112-pixels. There are variations in facial expressions (open/closed eyes, smiling/no-smiling), facial details (glasses), scale (up to 10%), and orientation (up to 20 degrees).

2. The Yale's database. It 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, w/glasses, 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-grey-level images (512x342-pixels) with slight variations of the head position (right into the camera, looking to the right, looking to the left, downwards, upwards), total images 300.

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 size of this set is 334 images.

5. Collection 1 database. A total number of 600 images (416x520), (432x544) which contain one person (different in each image) with variations mostly in expressions and uniform light conditions.

6. Collection 2 database. A total number of 2429 images (384x384-pixels and 384x360-pixels) which contain one person (different in each image) with variations in expressions, lighting conditions and background complexity.

7. Collection 3 database. A total number of 2578 images, which contain one person with variations in expressions, lighting conditions and background complexity, collected from the WWW.

4. EXPERIMENTAL RESULTS

The databases (6406 faces) described in the previous section were split in a set of training data and a set of testing data. Specifically, the first six Databases were used for the evaluation of the face detector (4228 faces) while the last collection of images was used to train the neural network (2578 faces). A set of 400 faces was used both for training and evaluation purposes.

For comparison reasons we built another face detector with the same NN topology but with some differences in the construction of the training database and the procedure. Analytically from each one of the 2578 hand labeled faces we produced 10 more face examples by random rotation about their center points (up to 9^o), scaling between 90% and 110%, translation one pixel, and mirroring. So the training database consisted of 25780 face patterns. The non-face patterns were collected with the same methodology. The face detector processed the original image without rotating it since we embedded this ability to the NN classifier.

DATABASES	NUMBER OF IMAGES	ACCURACY	REC.RATE (MISS IMAGES)
Olivetti	400	93.95%	97.75% (9)
Yale	165	93.67%	92.73% (12)
Bern	300	95.65%	99.00% (3)
FERET (Subset)	334	97.54%	100.00% (0)
Collection 1	600	92.67%	100.00% (0)
Collection 2	2429	94.50%	96.00% (97)
TOTAL	4228	-	97,14% (121)

Table 1: Detection results of the MLP network (M1).

DATABASES	NUMBER OF IMAGES	ACCURACY	REC.RATE (MISS IMAGES)
Olivetti	400	96.95%	99.00% (4)
Yale	165	93.52%	90.9% (15)
Bern	300	96.92%	99.33% (2)
FERET (Subset)	334	98.30%	99.70% (1)
Collection 1	600	93.30%	100.00% (0)
Collection 2	2429	95.70%	97.94% (50)
TOTAL	4228	-	98,297% (72)

Table 2: Detection results of the MLP network (M2).

The detection accuracy in all experiments was measured as follows. Each window area, identified by the neural network as a FR, was compared with the manually marked face area. In case that the common area of the FR and the manually marked area was greater than 0.7 of both areas union, a successful detection was assumed.

In table 1 the detection accuracy of the MLP method is shown for the case of increasing the number of the training images by rotating them in the range of (-9^0, $+9^0$), mirroring, and translating by one pixel the original face images (Method 1- M1). The face detector processed each original image without rotating it.

In table 2 the same measurements are shown for the case of performing rotation normalization in the training faces while the search space in the face detection process is increased by rotating the original image in the range of (-9^0, $+9^0$), using a step of 3^0 approximately (Method 2- M2).

5. CONCLUSIONS

The experimental results have given a face detection rate of 98.3% in a set of 4228 images for the proposed neural face detection system. This accuracy, carried out in a large face database, is the highest reported in the literature.

Currently our research is concentrated in a number of system restrictions. Specifically, the proposed method detects faces looking at 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 afterwards the proposed method can be used within these ROS for fine-tuning. Color segmentation methods can be used for the detection of ROS, reducing significantly the falsely accepted regions.

6. REFERENCES

[1] 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

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

[3] 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

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

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

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

[7] From the WWW (request: ackerman@iam.unibe.ch)

[8] 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, pp.39-51, Jan. 1998.

[9] 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, 1997