A face detector using Neural Networks and Discrete Wavelet Transforms

Inês Ap G Boaventura — IBILCE-UNESP — ines@ibilce.unesp.br
Valéria M Volpe — UNIRP — vmvolpe@unirpnet.com.br
André L V Sanches — UNIRP — andre@unirpnet.com.br
Adilson Gonzaga — EESC-USP — agonzaga@sc.usp.br

Abstract

This paper presents an approach for detecting human faces in digital images. Two different techniques are combined and used to accomplish the face detection task. First of all, Wavelet Transforms are applied in the input image in order to extract particular features that are detected by a locally connected Multilayer Perceptron network.

1. Introduction

Pattern recognition is a classic matter in computer vision that has been widely explored in many practical applications. The present paper describes a method for detecting specific features, based in the global image aspects, in order to detect human faces. This described system classifies images as face and non-face.

First of all, Wavelet Transforms are applied in the input image in order to extract particular features which are detected by a locally connected Multilayer Perceptron network [1],[2],[3]. The idea of locally connected neurons was inspired from Hubel and Wiesel's research on the locally sensitive neurons in the cat visual system [4].

This work aims to improve and simplify the feature extraction process proposed in [5], based in a convolutional neural network architecture which was initially proposed by LeCun et al [6].

The convolutional architecture is based on a highly complex process that automatically synthesizes the feature extractors from a training set of face and non-face patterns and classifies them accordingly.

The next section presents the proposed approach, results and discussion, as well as the conclusions on this study.

2. The solution for the problem

As already mentioned, the proposed solution is an intelligent mechanism which recognizes a given image as if being face or not. For that matter, two specific Wavelet functions, the Haar and Symlet, are used in order to extract features from a given image, which will then feed the Multilayer Perceptron locally connected input fields, responsible for their classification as being face or non-face. Before feeding the network, the input values are normalized to the range of -1 to 1.

Resulting from the Wavelet Transforms in a 32x32 pixel image, come eight 4x4 pixel images, being four of them processed by the Haar function, and the other four by the Symlet one. Each of these represents, respectively, the low-pass-low-pass (LL), low-pass-high-pass (LH), high-pass-low-pass (HL) and high-pass-high-pass (HH) filtering. In both cases, third level transformations wavelets were applied.

The proposed MLP consists of two layers of hyperbolic tangent neurons. The first layer, the hidden one, has eight neurons, where each one of them is locally connected to a specific resulting image from the Wavelet transforms. The second layer consists of only one neuron, fully-connected to the eight ones in the previous layer, which will output positive signal for face patterns and negative signal for the non-face ones. Figure 1 shows the details of the proposed architecture.

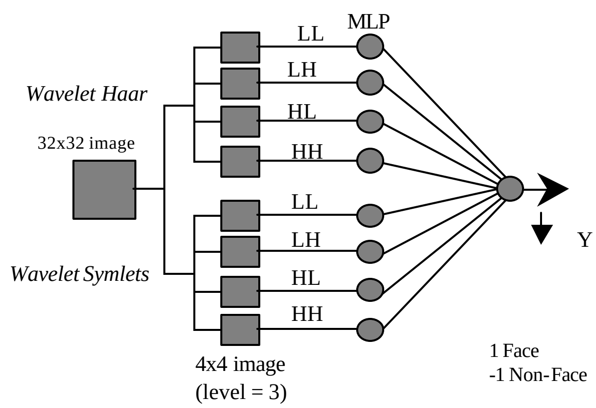

Figure 1. The proposed architecture.

The MLP was trained by using the simple backpropagation algorithm (no *momentum*), with learning rate of 0.001 and precision 0.5×10^{-6}. In order to satisfy the locally-connected input condition, the backpropagation algorithm was modified so that, during the weight adjustment in the hidden layer, only the input values related to its respective neuron were considered.

The used training set contained two hundred grayscale images of both face and non-face patterns. The face patterns were taken from the AR Database [7] with each of them being resized to 32x32 pixels using the nearest-neighbor interpolation and manually cropped to contain only features from eyes, nose and mouth. Non-face patterns were created by selecting images from random daily situations, such as landscapes and regular objects. It is also important to mention that the false positives resulting from previous network trainings were fed into the training set, so that the MLP could "learn from its own mistakes".

3. Results

To collect results, the MLP was trained for three times (T1, T2 and T3), with randomly initialized weights, in order to obtain the best generalization.

Thirty images were used for testing, sixteen of them being faces with a certain degree of complexity, and the other fourteen of them being non-face images.

The hardware used for the training was an Intel Pentium 4 processor with HyperThreading capability running at 3.06 GHz and 512 Mb of DDR RAM running at 333 MHz.

Table 1 shows the results of each training.

Table 1. Results

Training	Epoch	Total time	Correct answers
T1	2606	39.08s	100 %
T2	3179	48.03s	96,6 %
T3	1950	29.28s	96,6 %

It is important to mention that the images in the training set were taken in a controlled environment with neutral background color and position and intensity variable lighting.

4. Analysis of the results

Throughout the tests, the network presented a high error rate in non-face patterns. Thus, to improve its performance, the training set was increased with a greater number of non-face patterns, mainly taken from false positives of previous trainings, as mentioned before.

It was noticed that, in most false positive cases, the images contained occluded faces, such as people wearing glasses, which reflected the lights and causing confusion in the extraction of such features. In other cases, the faces were too bent to the sides, which caused misplacement of certain important features, such as in Figure 2.

Figure 2: Face bent to the side

On the other side, the network showed great efficiency in complex images, such as smiling faces, partially occluded images and faces with great lighting variation, as shown in Figure 3.

Figure 3. Glasses, lighting variation and smiling faces

5. Conclusions

By the analysis of the results, obtained from the referred test set, it can be said that they were well satisfying.

As a proposal for continuing this work, it is suggested the increasing of the training and test sets, inserting more partially occluded faces containing sunglasses, hats, beards and so on. It is also suggested the creation of an image rastering mechanism, in order to detect the spatial position of faces in real world images using the solution proposed in this paper.

6. References

[1] P. C. De Lima, "Wavelets: Uma Introdução", *Matemática Universitária*, 33, pp. 13-44, 2002.

[2] I. Daubechies, "Ten Lectures on Wavelets", *CBMS-NSF Regional Conference Series in Applied Mathematics*, Rutgers University and AT&T Bell Laboratories, volume 61, 1992

[3] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice-Hall, 1999.

[4] D. Hubel and T. Wiesel, “Receptive Fields, Binocular Interaction and Functional Architecture in the Cat’s Visual Cortex,”. J. Physiology, vol. 160, pp. 106-154, 1962.

[5] C. Garcia and M. Delakis, "Convolucional Face Finder: A Neural Architecture for Fast and Robust Face Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11),pp. 1408-1422, 2004.

[6] Y. LeCun, Y. Bengio and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", *Proceedings of the IEEE*, 86(11), pp.2278-2324, 1998.

[7] A. M. Martinez and R. Benavente, *The AR Face Database*, CVC Technical Report no. 24, June 1998.