A NEURAL NETWORK-BASED SYSTEM FOR FACE DETECTION IN LOW QUALITY WEB CAMERA IMAGES

Ioanna-Ourania Stathopoulou and George A. Tsihrintzis
Department of Informatics,University of Piraeus, Piraeus 185 34, Greece

Keywords: Face Detection, Multidimensional Signal Processing, Biometrics and Pattern Recognition, Neural Networks.

Abstract: The rapid and successful detection and localization of human faces in images is a prerequisite to a fully automated face image analysis system. In this paper, we present a neural network–based face detection system which arises from the outcome of a comparative study of two neural network models of different architecture and complexity. The fundamental difference in the construction of the two models lies in approaching the face detection problem either by seeking a general solution based on the full-face image or by composing the solution through the resolution of specific portions/characteristics of the face. The proposed system is based on the brightness contrasts between specific regions of the human face. We show that the second approach, even though more complicated, exhibits better performance in terms of detection and false-positive rates. We tested our system with low quality face images acquired with web cameras. The image test set includes both front and side view images of faces forming either a neutral or one of the "smile", "surprise", "disgust", "scream", "bored-sleepy", "angry", and "sad" expressions. The system achieved high face detection rates, regardless of facial expression or face view.

1 INTRODUCTION

Automated detection of human faces is one of the most difficult and important problems in the areas of pattern recognition and computer vision. Images that contain faces are instrumental in the development of more effective and friendlier methods for human-computer interaction. Vision-based human computer interaction methods assume that information about a user's identity, state and intent can be extracted from images, and that computers can then react accordingly. This information can also be used in a security control system to replace metal key, password, plastic card, or PIN number. Moreover, it can be used in criminology to uncover criminals. Finally, automated face detection is important in the area of *biometric authentication*, i.e., technologies that measure and analyze human physical and behavioural characteristics for authentication purposes.

Given an image, the goal of face detection is to determine whether there are any faces in the image and, if so, return the face location and extent. Such a problem is challenging because faces are not rigid and have a high degree of variability in size, shape, colour and texture. Furthermore, variations in pose, facial expression, image orientation and conditions add to the problem.

There have been developed three main approaches to the face detection problem, based on: (1) correlation templates, (2) deformable templates, and (3) image invariants, respectively. Correlation template-based approaches compute a difference measurement between one or more fixed target patterns and candidate image locations and the output is thresholded for matches. The use of deformable templates is similar in principle to the use of correlation templates, except that the latter are not rigid. In this approach, we try to find mathematical and geometrical patterns that depict particular regions of the face, fit the template to different parts of the images and threshold the output for matches. Finally, in image invariant-based approaches, the aim is to find structural features that exist even when pose, viewpoint and lighting conditions vary, and then use them to detect faces.

In the past, several systems have been developed that implement the above approaches. The system proposed by Colmenarez et al. is template-based and attempts to encode face images into a particular prototype (Colmenarez and Huang, 1997). Yang et al. (Yang and Huang 1994) and Lee et al. (Lee, Ham and Park, 1996) proposed knowledge-based systems

Stathopoulou I. and A. Tsihrintzis G. (2007).
A NEURAL NETWORK-BASED SYSTEM FOR FACE DETECTION IN LOW QUALITY WEB CAMERA IMAGES.
In *Proceedings of the Second International Conference on Signal Processing and Multimedia Applications*, pages 53-58
DOI: 10.5220/0002134100530058

that encode human knowledge of what constitutes a face. Leung et al. (Leung, Burl and Perona, 1995). applied a local feature detector to find faces in an image. Other systems (Rowley et al, 1997; Rowley et al, 1998; Yang and Huang 1994; Juell et al., 1996) use artificial neural networks to find faces. Lin et al. (Lin and Fan, 2001) proposed a system that searches for potential face regions, based on the triangle that form the eyes and the mouth. Sung et al. (Sung and Poggio, 1994) used metrics that measure the distance between the input image and the cluster of faces and non-faces. Lin Huang et al. (Huang and Shimizu, 2006) designed three detection experts which employ different feature representation schemes of local image and then use a polynomial neural network to determine whether or not there is a face in an image. Castrillon et al. (Castrillon et al., 2007) developed a system for real time detection of faces in video sequences by means of cue combination. S. Phimoltares et al. (Phimoltares, Lursinsap and Chamnongthai, 2007) developed a two-stage system, which first detects the faces from an original image by using Canny edge detection and their proposed average face templates and then uses a neural visual model (NVM) to recognize all possibilities of facial feature positions. Kadoury and Levine (Kadoury and Levine, 2007) proposed a novel technique which uses locally linear embedding (LLE) to determine a locally linear fit so that each data point can be represented by a linear combination of its closest neighbors and use this representation to train Support Vector Machines that detect faces.

Most of the aforementioned methods limit themselves to dealing with human faces in front view. There are several drawbacks in these approaches, such as: (1) They cannot detect a face, which is smaller than 50 * 50 pixels. (2) They cannot detect many faces (more than 3 faces) in complex backgrounds. (3) They cannot detect faces when we have images with defocus and noise problems. (4) They cannot all address the problem of partial occlusion of mouth or wearing sunglasses. (5) It is not easy to detect faces in side view. Although there are some researches that can solve two or three of these problems, there is still no system that can solve all of them.

In this paper, a new and efficient human face detection system is proposed that combines artificial neural networks and image invariants approaches. Specifically, in Section 2, we present our face detection algorithm. In Section 3, we discuss the structures of two neural networks, which have different performances. In Section 4, we present the face image data acquisition geometry and evaluate the performance of the two networks. Finally, we draw conclusion in Section 5 and point to future work in Section 6.

2 THE FACE DETECTION ALGORITHM

Our system uses face detection algorithms which fall within the third approach mentioned above. Specifically, we define certain image invariants and use them to detect faces by feeding them into an artificial neural network. These image invariants were found based on Sinha's (Sinha, P.; Yang and Ahuja, 2003) 14-by-16 pixel ratio template, summarized next.

2.1 The Sinha Template

The method proposed by P. Sinha (Sinha, P.; Yang and Ahuja, 2003) combines template matching and image invariant approaches. P. Sinha aimed at finding a model that would satisfactorily represent some basic relationships between the regions of a human face. More specifically, he found out that, while variations in illumination change the individual brightness of different parts of faces (such as eyes, cheeks, nose and forehead), the relative brightness of these parts remains unchanged. This relative brightness between facial parts is captured by an appropriate set of pairwise brighter - darker relationships between sub-regions of the face.

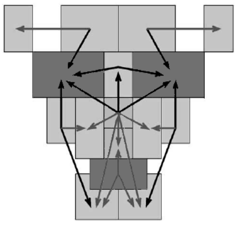


Figure 1: The Sinha Template.

The Sinha template is shown in Figure 1, where we observe 23 pair-wise relationships represented by arrows. The darker and brighter parts of the face are represented by darker and brighter shades of grey, respectively.

Our proposed face detection algorithm is built on this model. We pre-process a candidate image in order to enhance the relationships in the Sinha template and then feed the image into an artificial neural network to determine whether or not there is a face in the image.

2.2 The Algorithm

The main goal is to find the regions of the candidate image that contain human faces. Our system uses Artificial Neural Networks and operates in two modes: training the neural network and using it to detect faces in an image.

To train the neural network, we used a set of 285 images of faces and non-faces. We tried to find images of non-faces that are similar to human faces, so some of the non-face images contained dog, monkey and other animal "faces". These images where gathered from sources of the World Wide Web (Gender Classification (Databases)) and pre-processed before entered into the neural network.

To detect faces in a candidate image we apply a window, which scans the entire image, and pre-process each image region, the same way we pre-processed the images of the training set. Specifically, our algorithm works as follows:

1. We load the candidate image. It can be any 3-dimensional (color) image
2. We scan through the entire image with a 35-by-35 pixel window. The image region defined by the window constitutes the "window pattern" for our system, which will be tested to determine whether it contains a face. We increase the size of the window gradually, so as to cover all the possible sizes of a face in a candidate image.
3. We pre-process the "window pattern":
   - 3.1. We apply Histogram Equalization techniques to enhance the contrast within the "window pattern".
   - 3.2. We compute the eigenvectors of the image using the Principal Component Analysis and the Nystrom Algorithm ((Fowlkes, Belongie and Malik; Belongie, 2000; Shi and Malik, 2000; Image Segmentation using the Nystrom Method) to compute the normalized cuts.
   - 3.3. We compute three clusters of the image using the k-means algorithm and color each cluster with the average color.
   - 3.4. We convert the image from colored to grayscale (2D).
4. We resize the processed image into a dimension of 20-by-20 pixels and use it as input to the artificial neural network, which we present in the next section.

3 THE ARTIFICIAL NEURAL NETWORK STRUCTURES

To classify window patterns as "faces" or "non-faces", we developed two different artificial neural networks, which are presented next.

3.1 The First Artificial Neural Network

This network takes as input the entire window pattern and produces a two-dimensional vector output. The network consists of three hidden layers of thirty, ten and two neurons respectively, as in Figure 2. Its input ("window pattern") has dimension of 20-by-20 pixels. The neural network classifies the window pattern as "face" or "non-face". The output vector is of dimension 2-by-1 and equals to [1;0] if the window pattern represents a face or [0;1], else wise.

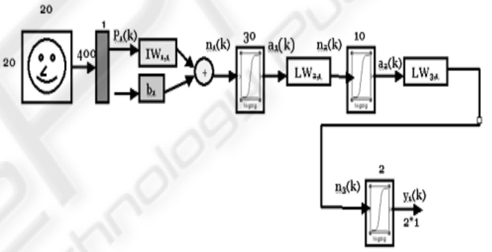


Figure 2: The First ANN's Structure.

3.2 The Second Artificial Neural Network

The second neural network has four hidden layers with one, four, four and two neurons, respectively. It is fed with the following input data: (1) the entire "window pattern" (20-by-20 pixels), (2) four parts of the "window pattern", each 10-by-10 pixels and (3) another four parts of the "window pattern", 5-by-20 pixels. Each of the three types of inputs is fed into different hidden layers of the network. The first, second, and third sets of inputs are fed into the first, second, and third hidden layer, respectively, while the output vector is the same as for the first network. Clearly, the first network consists of fewer hidden layers with more neurons and requires less input data compared to the second.

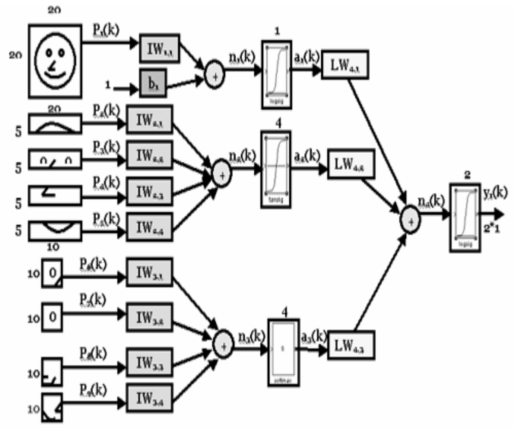


Figure 3: The Second ANN's Structure.

4 PERFORMANCE EVALUATION

4.1 Test Data Acquisition

The test image data were acquired with a three-camera system, as in Figure 4. Specifically, three identical cameras of 320-by-240 pixel resolution were placed with their optical axes on the same horizontal plane and successively separated by 30-degree angles. Subjects were asked to form facial expressions, which were photographed by the three cameras simultaneously.

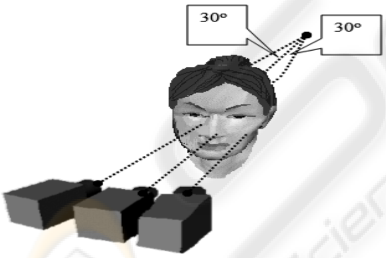


Figure 4: The geometry of the data acquisition setup.

The final dataset consisted of 250 different subjects, each forming the eight expressions: "neutral", "smile", "sad", "surprise", "angry", "disgust", "scream" and "bored-sleepy". Typical example images are shown in Table 3.

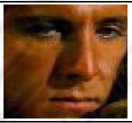
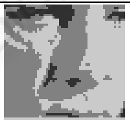
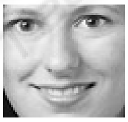
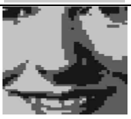
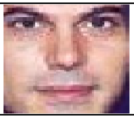
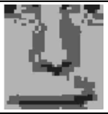
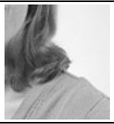
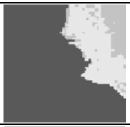
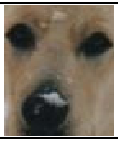
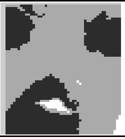
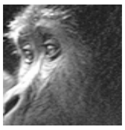
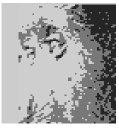
4.2 Performance Evaluation

To train the system networks, we used a common training set of 285 face and non-face images of relatively high quality. During the training process, the first and second networks reached an error rate of 10^{-1} and 10^{-10}, respectively.

We required that, for both networks, the output vector value be close to [1;0] when the window pattern represented a face and [0;1] otherwise. This means that the output vector corresponds to the degree of membership of the image in one of the two clusters: "face-image" and "non-face-image".

We tested the two networks with average quality face images first. Some results of the two neural networks can be seen in Table 1. The first network, even though it consisted of more neurons than the second one, did not detect faces in the images to a satisfactory degree, as did the second network. On the other hand, the execution speeds of two networks are comparable. Therefore, the second network was found superior in detecting faces in images and was decided to be used in our system.

Table 1: Results of the two neural networks for various images.

Input image	Pre-processed window pattern	First ANN's output	Second ANN's output
		[0.5; 0.5]	[0.947; 0.063]
		[0.6; 0.4]	[1 ; 0]
		[0.5; 0.5]	[0.9717; 0.0283]
		[0.5; 0.5]	[0 ; 1]
		[0.5; 0.5]	[0 ; 1]
		[0.5; 0.5]	[0 ; 1]

4.3 Evaluation on Low Quality face Images

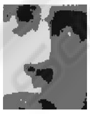
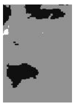
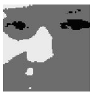
The aim of this evaluation was to examine whether the neural network was able to generalize satisfactorily with low quality face images and/or with the subject forming facial expressions and/or images acquired in side view. For this purpose, we tested our face detection system with 535 of the face images acquired with the three-camera system in Section 4. The dataset included a random selection of images in front and side view and images acquired from subjects forming one of the eight expressions: “neutral”, “smile”, “sad”, “surprise”, “angry”, “disgust”, “scream” and “bored-sleepy”.

Table 2: Face detection results for male and female face images.

Face Detection Results			
Genre	**Face Detected**	**Face no Detected**	**Success percentage rate**
Female	115	90	56,09%
Male	275	55	83,33%
Sum	390	145	**72,89%**

The system managed to detect face with 72,89% success rate. Errors (misses) occurred mostly because of overly bright illumination conditions which did not allow the extraction of facial features during k-means clustering.

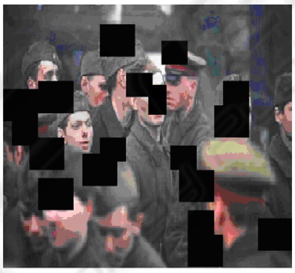
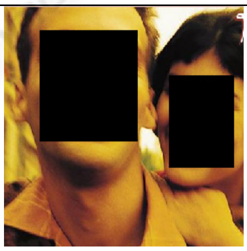
Table 3: Application of the proposed algorithm on low quality images.

Original Image	***Window Pattern***	***Pre-processed image***	***Network's Response***
			[1;0]
			[0,836; 0,164]
			[0,9531; 0,0469]

It was also observed that the detection of *female* faces was more difficult than the detection of *male* faces, possibly because facial features in female faces are not as tense as those in male faces. The results from these two groups are summarized in Table 2.

Some typical results of the face detection system are depicted in Tables 3 and 4. In Table 3, we show some results of applying the Face Detection Algorithm on the low quality web camera face images. In Table 4, we show the results of applying the algorithm to more complex images containing several rotated and partially occluded faces.

Table 4: Application of the proposed algorithm on two images.

	**6/8**
	**2/2**

5 SUMMARY AND CONCLUSIONS

In this work, we presented a neural network-based face detection system and tested it on low quality images acquired with a web camera set up. Although the neural network had been trained with a set of higher (digital camera) quality of images, it was able to generalize and detected the faces in images at a satisfactory rate. Occasionally, errors (e.g., failure to detect all faces in an image) occurred, especially with faces whose characteristics were not so clear. However, the system performance is expected to improve by widening the training set of the network.

6 FUTURE WORK

We plan to extend our work in the following three directions: (1) We will improve our system by retraining the neural network with a set that covers a wider range of poses and cases of low quality images. (2) We will examine face localization techniques localization to make the face detection task more rapid (3) Wee will integrate this system with a fully automated facial expression classification system, which we currently developing.

ACKNOWLEDGEMENTS

Support for this work was provided by the General Secretariat of Research and Technology, Greek Ministry of Development, under the auspices of the PENED-2003 basic research program.

REFERENCES

Belongie S. (2000). “Notes on Clustering Point-sets with Normalized Cuts”.

Castrillon, M. et al. (2007). ENCARA2: Real-time detection of multiple faces at different resolutions in video streams, Journal of Visual Communication and Image Representation –In Press.

Colmenarez, A. J. and Huang, T.S. (1997). “Face detection with information-based maximum discrimination”, Computer Vision and Pattern Recognition, 782–787.

Gender Classification (Databases): *http://ise0.stanford.edu/class/ee368a_proj00/project15/intro.html* *http://ise0.stanford.edu/class/ee368a_proj00/project15/append_a.html*

Fowlkes, C., Belongie, S. and Malik, J. ” Efficient Spatiotemporal Grouping Using the Nystrom Method”

Huang, Lin-Lin and Shimizu, Akinobu (2006). A multi-expert approach for robust face detection. Pattern Recognition 39(9): 1695-1703.

Image Segmentation using the Nystrom Method: *http://rick.ucsd.edu/~bleong/.*

Juell, P. and Marsh, R. (1996). “A hierarchical neural network for human face detection”, Pattern Recognition 29 (5), 781-787.

Kadoury, Samuel and Levine, Martin D. (2007). “Face detection in gray scale images using locally linear embeddings”, Computer Vision and Image Understanding, 105: 1–20.

Lee, S.Y., Ham, Y.K., Park, R.H. (1996). “Recognition of human front faces using knowledge-based feature extraction and neuro-fuzzy algorithm”, Pattern Recognition 29,1863-1876 (11) .

Leung, T.K., Burl, M.C. and Perona, P. (1995). “Finding faces in cluttered scenes using random labeled graph matching”, Fifth International Conference on Computer Vision, pages 637–644, Cambridge, Massachusetts, IEEE Computer Society Press.

Lin, C. and Fan, K. (2001). “Triangle-based approach to the detection of human face”, Pattern Recognition 34, 1271-1284.

Phimoltares, S., Lursinsap, C., Chamnongthai, K. (2007, May). “Face detection and facial feature localization without considering he appearance of image context”, Image and Vision Computing Volume 25 , Issue 5: 741-753.

Rowley, H.A., Baluja, S., and Kanade T. (1997). “Rotation Invariant Neural Network-Based Face Detection” CMU-CS-97-201.

Rowley, H.A., Baluja, S., and Kanade T. (1998). “Neural Network-based face detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1).

Shi, J. and Malik, J. (2000). “Normalized Cuts and Image Segmentation”, IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 22(8).

Sinha, P. “Object Recognition via image-invariants”.

Sung, K.K., Poggio, T. (1994). “Example-based learning for view-based human face detection”, *Proceedings on Image Understanding Workshop, Monterey, CA, 843-850.*

Yang, G. and Huang, T.S. (1994). “Human face detection in a complex background”, Pattern Recognition, 27(1):53–63 .

Yang, M.–H., Ahuja, N. (2003). “Face Detection and Gesture Recognition for Human- computer Interaction”, Kluver Academic Publishers.