Journal of ELECTRICAL ENGINEERING, VOL. 56, NO. 7-8, 2005, 195–199

A SYSTEM FOR LOCALIZATION OF HUMAN FACES IN IMAGES USING NEURAL NETWORKS

Marián Beszédeš — Miloš Oravec *

In this paper, a neural network-based face detection system inspired by the work of H. A. Rowley [3] is presented. Faces are detected in an unprocessed input image. The system processes and normalizes small windows extracted from the input image. Digital image processing techniques, such as normalization of size, position and rotation, improvement of light conditions and contrast are used here. A neural network is applied in two parts of the system. In the first one it detects rotation of the input window and afterwards it decides whether the window contains a face or not. In both cases a multilayer perceptron is used. The choice of the best topology and training method is discussed here, too. A face detection neural network uses the method of distributing the decision among multiple subnetworks. A special bootstrap algorithm is used to train the network. The result of the face detection system is in the form of a set containing locations of human faces.

K e y w o r d s: biometrics, face recognition, neural network, multilayer perceptron, bootstrap algorithm, digital image processing, histogram equalization

1 INTRODUCTION

In the recent past and now there is a growing interest in image processing algorithms that try to analyze the image content. These algorithms are used by applications which require automatic detection of human faces in images with a complex background. However, a face can have many variations in a scene. It can be anywhere in the image, have a variety of sizes, poses and illuminations.

Most of face detection algorithms can be divided into two major groups [1]:

The first one contains algorithms which detect faces by scanning an image for face-like pattern at many possible scales. Subimages of varying or constant sizes are extracted from the input image and classified according to the content as a face or a non-face region, using for example simple template matching [2], grey-level dependences [3], or eigenspace decomposition [1]. The approach based on template matching compares the shape of contours obtained by using edge operators with information about the shape of a face stored in the template. The approach based on grey-level dependences matches the distance of the intensity matrix obtained from a subimage with the intensity matrix of an "ideal" face. This distance measuring can be performed for example by a neural network. The concept of eigenspaces assumes that all face patterns can be well approximated in a low-dimensional subspace that is spanned by the eigenvectors ("eigenfaces") calculated from a set of test images. An image pattern is classified as a face if its distance to this face space is smaller than a certain threshold.

The second group contains algorithms based on image segmentation and is often used when colour information is available [4]. Due to the fact that colour is the most discriminating feature of a facial region, the first step of many face detection algorithms is pixel-based colour segmentation in order to detect skin-coloured regions. Subsequent classification based on the regions shape may fail if only parts of the face are detected or the face region is merged with a skin-coloured background. Thus, the performance depends mainly on the results of the initial segmentation.

Both groups of algorithms applied to the problem of face detection belong to the concept of biometrics [5]. Biometric recognition (or simply biometrics) refers to the automatic recognition of individuals based on their physiological and/or behavioural characteristics (such a characteristic is called a biometric characteristic, or biometric). A biometric system is a pattern recognition system that operates by acquiring biometric data from an individual, extracting a feature set from the acquired data, and comparing this feature set against the template set in the database

In this paper, an approach belonging to the first group of algorithms is described. It uses a set of neural networks [3]. At first, the neural network used here — multilayer perceptron — is briefly introduced. Then, the structure of the face localization system and description of its parts are described. The problem of the best topology, training method, and system parameters is discussed, too. The last section describes and discusses the obtained results.

2 MULTILAYER PERCEPTRON

The basic multilayer perceptron (MLP) building unit is a model of an artificial neuron. This unit computes the weighted sum of the inputs plus the threshold weight and passes this sum through the activation function (usually

* Department of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology, 812 19 Bratislava 1, Slovakia, beszedes@ktl.elf.stuba.sk, oravec@ktl.elf.stuba.sk

ISSN 1335-3632

sigmoid) [6, 7]:

$$v_j = \theta_j + \sum_{i=1}^{p} w_{ji} x_i = \sum_{i=1}^{p} w_{ji} x_i \tag{1}$$

$$y_j = \varphi_j(v_j) \tag{2}$$

where v_j is a linear combination of inputs $x_1, x_2, \ldots, x_p$ of neuron , $w_{j0} = \theta_j$ is the threshold weight connected to special input $x_0 = -1$, y_j is the output of neuron j and $\varphi(\cdot)$ is its activation function. Herein we use a special form of sigmoidal (non-constant, bounded, and monotone-increasing) activation function — logistic function

$$y_j = \frac{1}{1 + \exp(-v_j)}\,. \tag{3}$$

In a multilayer perceptron, the outputs of the units in one layer form the inputs to the next layer. The weights of the network are usually computed by training the network using the backpropagation (BP) algorithm.

A multilayer perceptron represents a nested sigmoidal scheme [1]. Its form for single output neuron is

$$F(x, w) = \varphi\Big(\sum_j w_{oj} \varphi\Big(\sum_k w_{jk} \varphi\Big(\ldots \varphi\Big(\sum_i w_{li} x_i\Big) \ldots\Big)\Big)\Big) \tag{4}$$

where $\varphi(\cdot)$ is a sigmoidal activation function, w_{oj} is the synaptic weight from neuron j in the last hidden layer to the single output neuron o, and so on for the other synaptic weights, x_i is the i-th element of the input vector $\boldsymbol{x}$. The weight vector $\boldsymbol{w}$ denotes the entire set of synaptic weights ordered by layer, then neurons in a layer, and then the number in a neuron.

In this work, also a bootstrap [8] procedure is used. Bootstrap techniques (also called resampling computation techniques) have introduced new advances in modelling and model evaluation. Using resampling methods, the information contained in one observed data set is extended to many typical generated data sets. Resampling allows the modeller to construct a series of new samples which are based on the original data set, and then to estimate the stability of the parameters. Properties such as convergence and asymptotic normality can be checked for any particular observed data set.

A bootstrap algorithm for face detection can be used as follows: It initializes the training set with examples of the face and non-face classes, then the neural network is trained, new images not containing faces are submitted to the network, false face detections are collected, and a part of the false detections is selected and incorporated to the non-face class of the training set.

3 STRUCTURE OF FACE LOCALIZATION SYSTEM

The structure of the system is shown in Fig. 1.

At first, an input image pyramid (a collection of images of reduced resolutions of the original image — meaning *eg* $1:2$, $1:4$, $1:8 \ldots 1:128$ (scale factor 2) is formed from a single input image by reducing it by a constant scale factor. Subimages are extracted by a window of constant size 20×20 pixels at every location and at every scale in the input image pyramid.

The steps of pre-processing are applied to every extracted subimage afterwards. The methods used for pre-processing are well known in image pre-processing [11]. Initially, pre-processing attempts to equalize the intensity values across the subimage [3]. It fits a function which varies linearly across the window to the intensity values inside the subimage. The linear function will approximate the overall brightness of the subimage and can be subtracted from the subimage to compensate for a variety of light conditions. Although the light conditions are corrected for, features are not set off. To reach contrast improvement, histogram equalization [11] is applied to the subimage to expand the range of intensities.

Pre-processed subimage serves as an input for the next part of the system which detects rotation of a possibly present face. This part consists of a neural network (see below) that *tries to predict the most face-like orientation of a given image region (even if it does not contain a face-like pattern). The suggested orientation is an output of this stage.*

With detected information, two steps of normalization can be done. The first one derotates the subimage around its centre according to the rotation information from the previous step. A classic image processing procedure is used here [11]. The second step applies a mask to the extracted subimage to eliminate useless background information from its corners. The mask shape is obtained empirically.

The normalized subimage is passed through the last part of the system that classifies the face occurrence probability [3]. This part is realized again by a neural network (see below). Real-valued output of the network indicates whether or not the subimage contains a face ($0 = 0\,\%$ of face presence, $1 = 100\,\%$ of face presence). If the system recognizes the subimage as a face, thus if the occurrence probability exceeds a selected threshold, then the subimage location is marked into the input image as a face area.

4 PARAMETERS OF THE SYSTEM AND TRAINING OF THE NEURAL NETWORKS

There are many parameters to be tuned in face localization system design and implementation. In this chapter, neural networks, training methods, and searching parameters will be described.

The neural network for rotation detection is a multilayer perceptron with a logistic activation function. The input layer of this network consists of 400 input units (20×20 pixel region is taken as an input). There are 30 hidden units formed into two hidden layers (15 units in

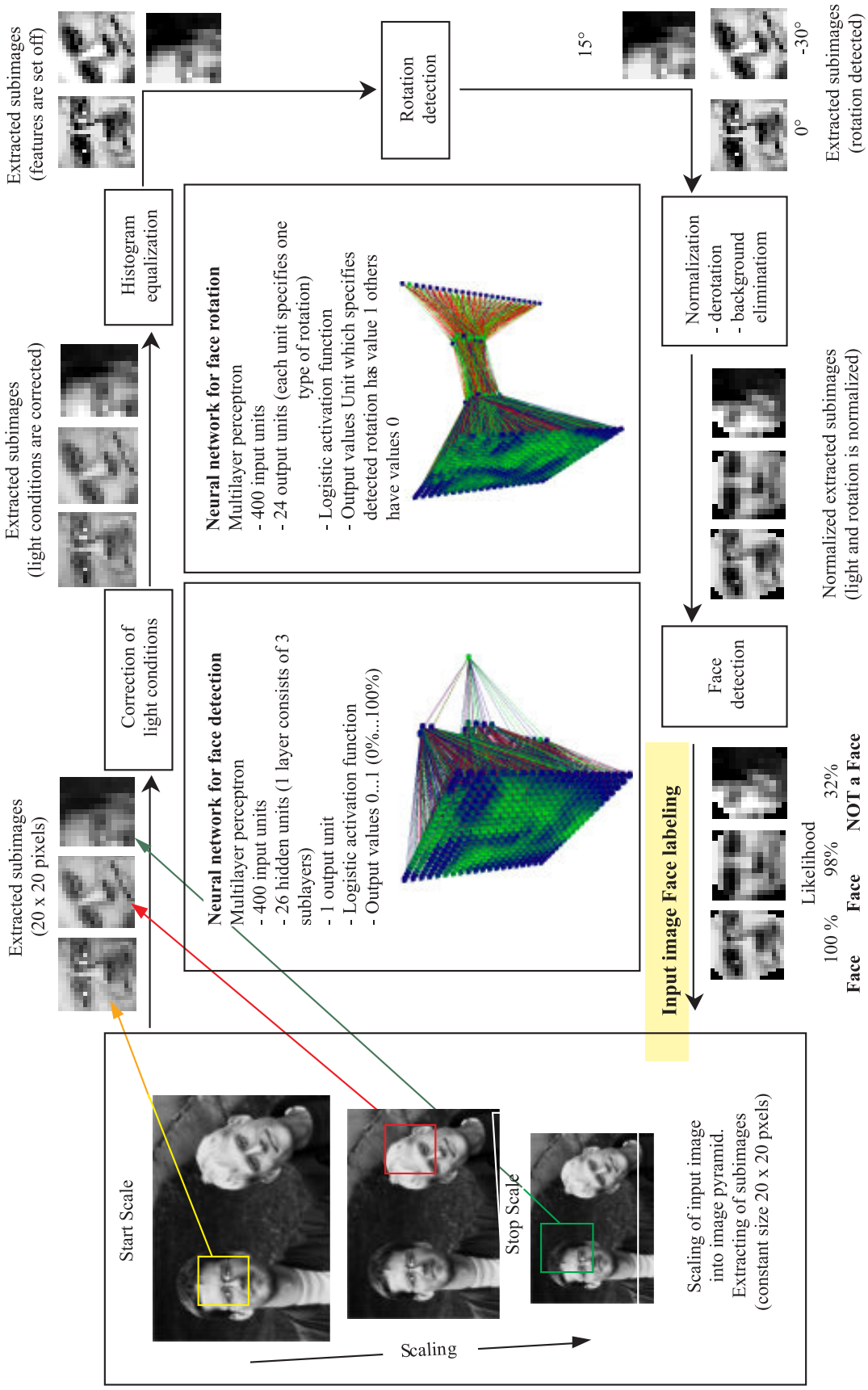

Fig. 1. Structure of face localization system

Fig. 2. Results of the proposed system

each layer). There are as many output units in the output layer as many rotations we want to detect. One output unit is used to determine one rotation value. We decided to detect 11 rotations with values $0°$, $\pm 15°$, $\pm 30°$, $\pm 45°$, $\pm 60°$, and $\pm 75°$. In an ideal case it means that if a face rotated by $60°$ is at the input, then the unit determined for this rotation has output 1 and other units have output 0.

The neural network for face detection is a multilayer perceptron with a logistic activation function again. The input layer has the same configuration as the previous one but the structure of the other parts is different. There is one hidden layer which consists of 26 hidden units. We can divide them into three groups, where each group of units analyses different size sub regions of input: 4 units which look at 10×10 pixel sub regions, 16 units which look at 5×5 pixel sub regions, 6 units which look at overlapping 20×5 horizontal stripes. Each of these types was chosen to allow the hidden unit to represent localized features that might be important for face detection. This neural network is trained by bootstrap algorithm with teacher presence to give output 1 when a face pattern is on the input, and 0 when a non-face pattern is on the input.

The training and testing set for rotation detection contained 3850 samples generated by rotation of 350 different faces (11 rotations — $0°$, $\pm 15°$, $\pm 30°$, $\pm 45°$, $\pm 60°$, $75°$). This neural network was trained by a back propagation algorithm with training set shuffling. The weights were randomly initialized by the values from the range $\langle -1, 1 \rangle$ and after many experiments we decided to set the learning factor eta to a value of 0.2. The network was trained to achieve the best generalization properties [7].

The basic training and testing set for face detection contained 250 face and 250 randomly generated non-face samples. The training and testing sets were extended according to the boosting method by 2000 and 2500 non-face samples, respectively. These samples were obtained as false face detections from testing of a neural network trained on the basic sets on images containing no faces.

The result of the training of both networks is strongly dependent on the initialization of weights. We tried to initialize each of the networks in 10 different ways and we took only the networks with the best generalization ability. We tried to change the number of hidden layers in both neural networks. The best generalization ability was found for the network with 2 hidden layers for the rotation detection network, and with 1 hidden layer for face detection.

The correctness of results and the speed of localization process depended on system parameters such as the size of input image, factor of input image scaling or value of extraction window shift. The best value for input image scaling is from 1.02 to 1.07. When the value of the extraction window shift is too small, then detection is too slow, but there is no risk of loss of any face occurrence. On the other hand, when the value of extraction window shift is too big, then detection is fast but face occurrences are lost. We found that values 2 and 3 are the best for our implementation.

5 RESULTS AND FUTURE DIRECTIONS

Figures 2 and 3 show the results of the proposed face localization system. It can be seen that every face was found more then once. Our system succeeded in 90100 percent, when we consider the number of faces in the input picture. For instance, in Fig. 3 one can see a test where almost 2 million of windows were extracted from the input image. Only 18 from this amount were not classified correctly. Only three faces from the total number 38 were not found. But there is still a rather big count of false detections even when the bootstrap algorithm is used. The solution could be based on the fact that every face was detected many times and wrong detections

Fig. 3. Other results of the proposed system

are usually single. There should be another part of the system which combines overlapping multidetection and prevents single detection. Then, better results should be expected. Segmentation and region processing [9, 10] is a promising way for face localization algorithms. Another way how to improve the speed of the system is to use a colour scheme of the human face to detect just regions of possibly present human faces. The processing described above should be used only in these regions.

6 CONCLUSION

A neural network-based face detection system with description of all critical parameters was presented. Digital image processing techniques, such as normalization of size, position and rotation, improvement of light conditions and contrast were used here. Neural networks were applied in two parts of the system. In the first one it detected rotation of the input window and in the other one it decided whether the window contains a face or not. The presented scheme is appropriate for today's face recognition systems.

References

[1] MENSER, B.—BRUNIG, M.: Locating Human Faces in Colour Images with Complex Background, In Proc. IEEE Int. Symposium on Intelligent Signal Processing and Communication Systems ISPACS '99, pp. 533–536, Phuket, Thailand, December 1999.

[2] JEBARA, T. S.: 3D Pose Estimation and Normalization For Face Recognition, Montreal, McGill University, 1996.

[3] ROWLEY, H. A.: Neural Network-Based Face Detection, Pittsburgh, Carnegie Mellon University, 1999, PhD thesis.

[4] MENSER, B.—WIEN, M.: Segmentation and Tracking of Facial Regions in Colour Image Sequences, In Proc. SPIE Visual Communications and Image Processing 2000, volume 4067, pp. 731–740, Perth, Australia, June 2000.

[5] ANIL K. JAIN—ARUN ROSS—SALIL PRABHAKAR: An Introduction to Biometric Recognition, IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video-Based Biometrics **14** No. 1 (2004), 4–20.

[6] HAYKIN, S.: Neural Networks — A Comprehensive Foundation, Macmillan College Publishing Company, New York, 1994.

[7] ORAVEC, M.—POLEC, J.—MARCHEVSKÝ, S. *et al*: Neural Networks for Digital Signal Processing, Faber Bratislava, 1998. (in Slovak)

[8] KALLEL, R.—COTTRELL, M.—VIGNERON, V.: Bootstrap for Neural Model Selection, in Proc. of ESANN'2000 – European Symposium on Artificial Neural Networks, Bruges (Belgium), 26-28 April 2000, pp. 61–68.

[9] PAVLOVIČOVÁ, J.—POLEC, J., NEDIELKA, R.: Segmentation for Face Oriented Image Compression, Proc. of IWSSIP 2000, Maribor, Slovenia 7-9 June, 2000, pp. 131–134.

[10] POLEC, J.—PAVLOVIČOVÁ, J.—VANYA, R.—KARLUBÍKOVÁ, T.—MASLÍK, Ľ.—PARTYK, M.: New Scheme for Region Approximation and Coding with Shape Independent Transform, Proc. of ISPRS, Comission III, Vol. 34, Part 3B "Photogrammetric Computer Vision", Graz, 2002, pp. B214–217.

[11] GONZALES, R. C.—WOODS, R. E.: Digital Image Processing, Addison-Wesey Publishing Comp., 1992.

Received 7 September 2004

Marián Beszédeš (Ing) received his Ing degree from the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in 2003. He is currently a PhD student at the Department of Telecommunications. His current research interests include neural networks and signal processing.

Miloš Oravec (Doc, Dr, Ing) received his Ing, Dr, and Doc degrees from the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in 1990, 1996, and 2002, respectively. Now, he is Associate Professor at the Department of Telecommunications, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology. His current research interests include neural networks, signal processing and telecommunication networks. He is a member of IEEE and IEE.