Journal of Software Engineering and Applications, **2011, 4, 682-687**
doi:10.4236/jsea.2011.412080 Published Online December 2011 (http://www.SciRP.org/journal/jsea)

Face Detection and Localization in Color Images: An Efficient Neural Approach

Samy Sadek[1], Ayoub Al-Hamadi[1], Bernd Michaelis[1], Usama Sayed[2]

[1]Institute for Electronics, Signal Processing and Communications (IESK), Otto-von-Guericke-University Magdeburg, Magdeburg, Germany; [2]Department of Electrical Engineering, Assiut University, Assiut, Egypt.
Email: samy.bakheet@ovgu.de, samy.technik@yahoo.de

Received October 17th, 2011; revised November 29th, 2011; accepted December 11th, 2011.

ABSTRACT

Automatic face detection and localization is a key problem in many computer vision tasks. In this paper, a simple yet effective approach for detecting and locating human faces in color images is proposed. The contribution of this paper is twofold. First, a particular reference to face detection techniques along with a background to neural networks is given. Second, and maybe most importantly, an adaptive cubic-spline neural network is designed to be used to detect and locate human faces in uncontrolled environments. The experimental results conducted on our test set show the effectiveness of the proposed approach and it can compare favorably with other state-of-the-art approaches in the literature.

Keywords*: *Human Face Detection and Localization, Spline Activation Function, Color Moments, Human-Computer Interaction

1. Introduction

Automatic face detection and localization is an active area of research spanning several disciplines in computer vision and pattern classification and has many application potentials, yet it still presents one of the most challenging computer vision problems. For instance, mugshot matching, user verification and access control, enhanced human-computer interaction, and crowd surveillance all are becoming possible if an effective face detection system could be implemented [1]. There are two fundamental face detection techniques: content-based methods and color-based methods. Content-based methods try to identify features in a human face. Most content-based methods were developed for grayscale images to avoid the complexity of combining the features detected in the RGB color space. A method developed by Yow and Cipolla [2] elongates the image in the horizontal direction and identifies thin horizontal features, such as the eyes and mouth. Cootes and Taylor [3] develop a technique that matches features to a model face using statistical methods. Leung *et al*. [4] present a similar method that matches features to a model face, except they used a graph matching algorithm to compare detected features to the model. In [5], Rowley *et al*. develop a front view face detection system that uses neural networks to pick out features. Instead of using neural networks, Sung and Poggio [6] develop an example-based learning technique, while Colmenzrez and Huang [7] use a probabilistic visual learning system. A survey of content-based techniques for general image retrieval can be found in [8].

Unfortunately, content-based techniques are very complex and expensive computationally. Also, if the face is rotated or partially obscured, the technique has to incurporate other techniques to solve the image registration and occlusion problems. In the other hand, color-based methods are based on calculating histograms of the color values and then develop a chroma chart to identify the probability that a particular range of pixel values represent human flesh. It has been found that the effectiveness of the method depends highly on the color space used. Chroma charts have been developed for the standard RGB color space [9], the YIQ color space [10], the HSV color space [11,12], and the LUV space [13]. The implementation of color-based techniques is fairly simple and, after the system has learned a chroma chart, the processing is very efficient. Also, the methods handle color images in a more straightforward manner than the content-based methods. However, as [14] describes, color-based techniques have several drawbacks. These disadvantages include information loss due to quantization, the strong dependence on the color space, and erroneous retrieval in the presence of gamma nonlinearity. The most significant drawback, however, is that a technique based

solely on a color histogram ignores all spatial information in the image. That is, color histograms catalog the global distribution of colors, but do not tell how the colors are arranged to form shapes and features. Despite these disadvantages, color histograms are very popular due to their simplicity and ease of calculation.

The remainder of the paper proceeds as follows. Section 2 outlines the neural model (*i.e*. cubic-spline neural network) used as classifier with the proposed approach. In Section 3, we describe the proposed method that is based on YES histograms and color moments. In Section 4, experimental results are reported. Finally, a few concluding remarks and suggestions for possible future extensions are given in Section 5.

2. Cubic-Spline Neural Networks

Artificial Neural networks (ANNs) are very likely to be the future of computing. A neural network is a powerful data modeling tool that is able to capture and represent complex input/output relationships. The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain. A graphical representation of the neural model is shown in **Figure 1**. The ANN learns via a process called "training". With training, the input data is repeatedly presented to the neural network. With each presentation the output of the neural network is compared to the desired output and an error is computed. This error is then fed to the neural network and used to adjust the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output.

To produce an output closer to the desired output, the neurons of network employ a non-linear function, so-called activation function which is usually a non-linear monotonic function and generally based upon the sigmoidal function. The activation function simulates the correlation between the action potential of the inputs and the output of the neuron. In this work we employ an adaptive activation function for the hidden neurons out of a pool of standard functions called cubic-spline function to increase flexibility [15]. The neural model employing this type of activation function is called Cubic-spline Neural Network (CSNN). Mathematically, the cubic-spline activation function is defined by

$$S(x) = s_k = \sum_{i=1}^{3} s_{k,i} (x - x_k)^i \qquad (1)$$

$$\forall x \in [x_k, x_{k+1}], k = 1, 2, \cdots, n-1.$$

where $s_{k,i}$ are the coefficients of the cubic-spline function. Further details on this model can be found in [16].

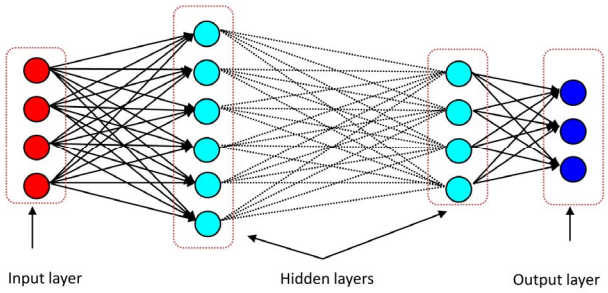

Figure 1. Block diagram of an ANN architecture.

3. Proposed Methodology

This section is to discuss the proposed methodology for real-time face detection and localization. **Figure 2** is a simplified block diagram illustrating the main components of the proposed architecture, and how they interact with each other in order to achieve effective functionality of the whole approach. As illustrated in the block diagram, the proposed approach generally consists of two parts, each carries out a specific tasks. The first part performs face detection task, while the second one performs face localization task. Each of these two tasks can be described briefly below.

Face Detection

To achieve this task, the proposed approach tries to discriminate between two classes of images (*i.e*., "face" class and "non-face" class). It is noted that training a neural model for the face detection task is challenging because of the difficulty in characterizing prototypical "non-face" images. It is easy to get a representative sample of images which contain faces, but it is much difficult to get a representative sample of those which do not. A simple procedure for this task works as follows: At first, the feature vector x which consists of information (YES histograms and color moments) derived from a given image is fed into the designed adaptive SNN. Then, the output *y* will represent the probability that the image contains a human face. Formally, the output *y* for a given image can be interpreted as:

$$y = p(\mathrm{x}) = \begin{cases} > 0.5, & \text{face found} \\ o.w, & \text{face not found} \end{cases} \qquad (2)$$

Face Localization

In terms of the methodology for face localization task,

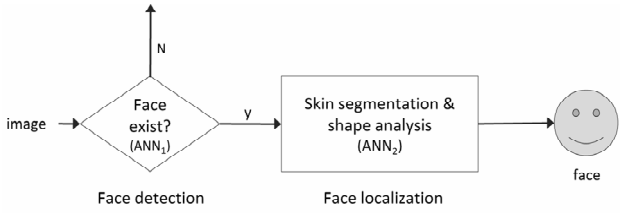

Figure 2. Main structure of the proposed approach.

the approach attempts to identify the location of a face in a given image. The ultimate goal of this task is finding an object in an image as a face candidate that its shape resembles the shape of a face. Faces can be characterized by elliptical shape and an ellipse can approximate the shape of a face. In order to perform face localization task, the proposed approach carries out two subtasks: a human skin segmentation to identify possible regions corresponding to human faces; and shape analysis to separate isolated human faces from initial segmentation results and then identify the location of each human face in the image. In the following subsections, we discuss the different modules that implement the baseline architecture aforementioned in **Figure 2**, with a particular focus on the feature extraction module.

3.1. Preprocessing

For later successful feature extraction and classification, it is important to preprocess all video sequences to remove noisy, erroneous, and incomplete data, and to prepare the representative features that are suitable for knowledge generation. To wipe off noise and weaken image distortion, all frames of each action snippet are smoothed by Gaussian convolution with a kernel of size 3×3 and variance $\sigma = 0.5$.

3.2. Feature Extraction

Feature extraction is indeed the core of any recognition system, but is also the most challenging and timeconsuming part. Further it was stated that the overall performance of the recognition system relies heavily on the feature extraction than the classification part. In particular, real-time feature extraction is a key component for any action recognition system that claims to be truly real-time. Many varieties of visual features can be used for face detection and localization. In this work, the features that have been considered are derived from the difference images that primarily describe the shape of the moving human body parts. Such features represent a fundamental source of information regarding the interpretation of a specific human action. Furthermore the information of motion can be also extracted by following the trajectory of the motion centroid. The extracted features are primarily based on computing the moments of the difference images to specify the type of motion of a given action. Therefore the basic features are defined as:

YES Histogram

RGB is the natural color space to work in, since most color images are encoded in this space. Although the RGB histogram may yield some positive results in many color based image retrieval or classification systems [17], it is still not a satisfactory face detection system. The transformation from the standard RGB color space to the YES color space is given by the following matrix equation:

$$\begin{pmatrix} \mathrm{Y} \\ \mathrm{E} \\ \mathrm{S} \end{pmatrix} = \begin{pmatrix} 0.253 & 0.684 & 0.063 \\ 0.5 & -0.5 & 0.0 \\ 0.25 & 0.25 & -0.5 \end{pmatrix} \begin{pmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{pmatrix} \tag{3}$$

It is noted that the Y component picks out the edges of the image, while the E and S components encode the color intensities. The Y histogram may, in some sense, provide the neural network with spatial information, rather than just color intensities. The errors in the RGB histogram approach may have been due in part to the similarity of the three R, G, and B histograms. The YES approach seemed to have resulted in histograms that with greater variation. In this manner, appending the three histograms along with color moments as a vector provides the network with more information.

Color Moments

Color moments have been successfully used in many color-based image retrieval or classification systems [18-21], especially when the image contains just the object. The first order (mean), the second (variance) and the third order (skewness) color moments have been proved to be efficient and effective in representing color distributions of images. Mathematically, the first three moments can be defined as:

$$\mu_k = \frac{1}{\lambda}\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij}^{k}$$

$$\sigma_k = \sqrt{\frac{1}{\lambda}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(p_{ij}^{k}-\mu_k\right)^2}$$

$$s_k = \sqrt[3]{\frac{1}{\lambda}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(p_{ij}^{k}-\mu_k\right)^3}$$

where p_{ij}^{k} is the value of the *k-th* color component of the image *ij-th* pixel and $\lambda = mn$ where *m* and *n* are the height and the width of the image, respectively.

3.3. Feature Classification

In this section, face detection task is modeled as a simple two-class classification task, and the goal is to assign a class to a given image. There are various supervised learning algorithms by which a face detection can be trained. The neural classifier aforementioned in Section 2 is used for the current classification task due to its outstanding generalization capability and reputation of a highly accurate paradigm. The basic model of the ANN classifier that we used is an MLP network with multiple hidden layers with 20 neurons each, which is most similar to the classical network structures but with improving in the hidden-unit adaptive activation functions (*i.e*. the hyperbolic-tangent function). Before the training phase, the

classifier begins with random weights at the connections between the neurons. The learning procedure followed by the ANN classifier is similar to the well-known back-propagation procedure [22,23]. In our approach, two classes of images are created. During the learning stage, the ANN classifier is trained using the features extracted from the images in the training set. The 24-bin YES histograms (8-bin for each component) representing the color features are first transformed into plain vectors, and then fused with the image-moments features. All feature vectors are finally fed into the ANN classifier to distinguish the image classes. After the learning stage is finished, the system is able to detect and identify unseen image. In fact, the classifier produces a real value between 0 and 1 which can easily be binarized by using a predetermined threshold.

4. Experimental Results

In this section, the experiments conducted to assess the performance of the proposed approach are described and some of their results are presented. In order to prepare the experiments and to provide an unbiased estimation of the generalization abilities of the classification process, the images in our dataset were partitioned into two independent subsets, *i.e*. a training set and a test set. More specifically, a set of images (50% of all images) were used for training and other image (the remaining 50%) were set aside as a test set. An MLP network with 33 input, 20 hidden and 1 output neurons was trained on the training set, while the evaluation of the detection performance was performed on the test set. The first half of the training set were labeled as face images, while the second half were labeled as non-face. The face-labeled images were chosen to represent a variety of ages, genders, and skin tones. The other non-face images represent different objects randomly collected from internet sites. Some of the non-face images were chosen to "fool" the neural classifier. For instance, some of these images contained flesh tones or facial features. After the training process, the neural classifier could correctly classify all training images. The detection results obtained on the test set are outlined in **Table 1** and depicted graphically in terms of true positive (TP) and false positive (FP) vs the number of hidden layers of the network in **Figure 3**. It may not be irrelevant to mention here that some of non-face images in the test set contained skin tones that were not represented in the training set (see **Figure 4**). These issues were a big challenge for the proposed approach to identify these images correctly. **Figure 5** shows some results obtained with the proposed method when applied on "multi-face" images in the test set. These evaluation results demonstrate that the proposed approach not only can detect human faces, but also can accurately localize them in multi-face images.

Table 1. Accuracy performance vs. no. of hidden neurons.

No. of hidden layers:	1	2	3	4	5
Average true positive (TP)	0.85	0.92	0.95	0.98	0.97
Average false positive (FP)	0.18	0.12	0.010	0.09	0.11
RMS error			$<1.0\times10^{-10}$		

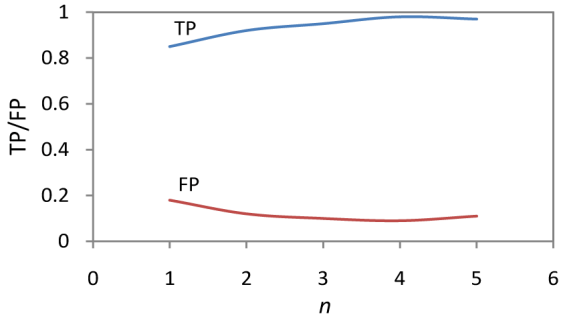

Figure 3. Detection performance in terms of true positive (TP) and false positive (FP) vs the number of hidden layers of the network.

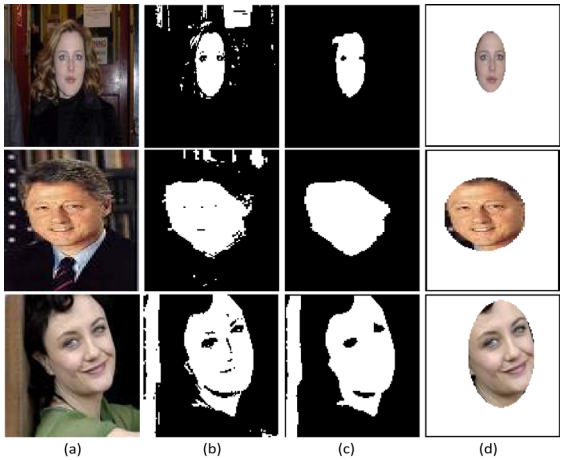

Figure 4. Some results for "face" image localization: (a) Source image; (b) Skin-colored regions; (c) Filtered image; and (d) Localized face image.

Figure 5. Results for "multi-face" image localization.

5. Conclusions and Future Work

In this paper we have presented a computationally efficient approach for real-time face detection and localization in color images using a finite set of low-level features directly derived from the input image. The obtained results showed that using YES histograms and color moments to detect and localize face is a promising approach. The key advantage of the proposed approach is that the training process takes a trivial time to complete. Furthermore the approach can locate multiple faces with encouraging results that enable the proposed approach to compare favorably with other state-of-the-art approaches in terms of detection and false-positive rates. Additionally, the process of locating multiple faces in image does not enlarge time-consuming, so that the approach can offer timing guarantees to real-time applications. However, it would be advantageous to explore the empirical validation of the approach on more complex large benchmark video datasets presenting many technical challenges in data handling such as object articulation, occlusion, and significant background clutter. These issues are of great interest and could be more complex, so that we plan to address them thoroughly in our future work.

6. Acknowledgements

This work is supported by Transregional Collaborative Research Centre SFB/TRR 62 “Companion-Technology for Cognitive Technical Systems” funded by DFG, and BMBF Bernstein-Group (FKZ: 01GQ0702).

REFERENCES

[1] K. Sung and T. Poggio, “Example-Based Learning for View-Based Human Face Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, 1998, pp. 39-51. doi:10.1109/34.655648

[2] K. Yow and R. Cipolla, “Feature-Based Human Face Detection,” *Image and Vision Computing*, Vol. 2, No. 15, 1997, pp. 713-735. doi:10.1016/S0262-8856(97)00003-6

[3] T. Cootes and C. Taylor, “Locating Faces Using Statistical Feature Detectors,” *Proceeding of the Second International Conference on Automatic Face and Gesture Recognition*, Killington, 14-16 October 1996, pp. 640-645. doi:10.1109/AFGR.1996.557265

[4] T. Leung, M. Burl and P. Perona, “Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching,” *Proceedings of the Fifth International Conference on Computer Vision*, Cambridge, 20-23 June 1995, pp. 637-644.

[5] H. Rowley, S. Bluja and T. Kanade, “Neural Network-Based Face Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, 1998, pp. 23-38. doi:10.1109/34.655647

[6] K. Sung and T. Poggio, “Example-Based Learning for Viewbased Human Face Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, 1998, pp. 39-51. doi:10.1109/34.655648

[7] A. Colmenarez and T. Huang, “Face Detection with Information-Based Maximum Discrimination,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 17-19 June 1997, pp. 278-287.

[8] M. D. Mariscoi, L. Cinque and S. Levialdi, “Indexing Pictorial Documents by Their Content: A Survey of Current Techniques,” *Image and Vision Computing*, Vol. 15, No. 2, 1997, pp. 119-141. doi:10.1016/S0262-8856(96)01114-6

[9] B. Schiele and A. Waibel, “Gaze Tracking Based on Facecolor,” *International Workshop on Face and Gesture Recognition*, Zurich, 1995.

[10] Y. Dai and Y. Nakano, “Face-Texture Model Based on Sgld and Its Applications in Face Detection in a Color Scene,” *Pattern Recognition*, Vol. 29, No. 6, 1996, pp. 1007-1017. doi:10.1016/0031-3203(95)00139-5

[11] Q. Chen, H. Wu and M. Yachida, “Face Detection by Fuzzy Pattern Matching,” *Proceedings of the Fifth International Conference on Computer Vision*, Cambridge, 20-23 June 1995, pp. 591-596.

[12] J. Cai and A. Goshtasby, “Detecting Human Faces in Color Images,” *Image and Vision Computing*, Vol. 18, 1999, pp. 63-75. doi:10.1016/S0262-8856(99)00006-2

[13] Y. Miyake, H. Saitoh, H. Yaguchi and N. Tsukada, “Facial Pattern Detection and Color Correction from Television Picture and Newspaper Printing,” *Journal of Imaging Technology*, Vol. 16, No. 5, 1990, pp. 165-169.

[14] D. Androutsos, K. N. Plataniotois and A. N. Venetsanopoulos, “A Novel Vector-Based Approach to Color Image Retrieval Using a Vector Angular-Based Distance Measure,” *Computer Vision and Image Understanding*, Vol. 75, No. 1-2, 1999, pp. 46-58. doi:10.1006/cviu.1999.0767

[15] S. Sadek, A. Al-Hamadi, B. Michaelis and U. Sayed, “Image Retrieval Using Cubic Spline Neural Networks,” *International Journal of Video & Image Processing and Network Security* (*IJIPNS*), Vol. 9, No. 10, 2009, pp. 17-22.

[16] S. Sadek, A. Al-Hamadi, B. Michaelis and U. Sayed, “Cubic-Spline Neural Network-Based System for Image Retrieval,” *Proceedings of Sixth International IEEE Conference on Image Processing* (*ICIP*’09), Cairo, 7-11 November 2009, pp. 273-276.

[17] S. Sadek, A. Al-Hamadi, B. Michaelis and U. Sayed, “A Robust Neural System for Objectionable Image Recognition,” *Proceedings of Second International Conference on Machine Vision* (*ICMV*2009), Dubai, 28-30 December 2009, pp. 32-36.

[18] S. Sadek, A. Al-Hamadi, B. Michaelis and U. Sayed, “A New Method for Image Classification Based on Multi-Level Neural Networks,” *Proceedings of International Conference on Signal and Image Processing* (*IC-SIP*2009), Amsterdam, 29 July-1 August 2009, pp. 197-200.

[19] B. Si, W. Gao, H. Lu and W. Zeng, “An Image Retrieval Method Based Regions of Interest,” *High Technology Le-*

tters, Vol. 13, No. 5, 2003, pp. 13-18.

[20] S. Sadek, A. Al-Hamadi, B. Michaelis and U. Sayed, "An Image Classification Approach Using Multilevel Neural Networks," *Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems* (*ICIS*'09), Shanghai, 17-20 September 2009, pp. 180-183.

[21] S. Sadek, A. Al-Hamadi, B. Michaelis and U. Sayed, "An Efficient Approach for Region-Based Image Classification and Retrieval," *Communications in Computer and Information Science*, Vol. 61, 2009, pp. 56-64. doi:10.1007/978-3-642-10546-3_8

[22] D. Rumelhart, G. Hinton and R. Williams, "Learning Internal Representation by Error Propagation," *Parallel Distributed Processing*: *Explorations in the Microstructures of Cognition*, Vol. 1, MIT Press, Cambridge, 1986.

[23] C. Bishop, "Neural Networks for Pattern Recognition," Oxford University Press, Oxford, 1995.

