Improved Neural Network-based Face Detection Method using Color Images

Yuriy Kurylyak[1], Ihor Paliy[1], Anatoly Sachenko[1], Kurosh Madani[2] and Amine Chohra[2]

[1] Research Institute of Intelligent Computer Systems
Ternopil National Economic University
3 Peremoga Square, 46004, Ternopil, Ukraine
{as, ipl, yuk}@tanet.edu.te.ua

[2] Images, Signals and Intelligent Systems Laboratory (LISSI / EA 3956)
PARIS XII University, Senart-FB Institute of Technology
Av. Pierre Point, Bat. A, F-77127, Lieusaint, France
{madani, chohra}@univ-paris 12.fr

Abstract. The paper describes some face detection algorithms using skin color segmentation, Haar-like features and neural networks. The segmentation using skin color labels promising input image areas that may contain faces. The usage of Haar-like features allows fast rejection of the majority of background. Then, the ensemble of retinally connected neural networks performs the final classification of the rest image windows using improved face search strategy across scale and position. The proposed search strategy applies inverse image scale pyramid, adaptive scanning step and window acceptance to decrease the number of windows which should be processed by the classifier.

1 Introduction

Human face detection (FD) is a very important quick-developing research area which has a wide range of applications, like face recognition, video-conference, content-based image retrieval, video-surveillance, etc. FD is also a challenging task because of facial variability in scale, location, orientation and pose. Many different FD approaches have been proposed in the last years: knowledge-based, invariant feature-based, template matching, and appearance-based [1]. The earlier methods are based on human-coded rules or facial features which are invariant to pose and orientation change, difficulty handle cluttered scenes with complex background and detect a lot of false positives [1]. Some facial features like a skin color may be used to select face candidate regions which extremely reduces the search area. Then these regions may be processed by more complex and accurate classifier. The simplest skin color segmentation method is pixel-based skin color detection with explicitly defined skin cluster boundaries in some color-spaces [2]. Applying of some Haar-like features also reduces the search space [3].

More recent FD methods from appearance-based group show excellent results on benchmark test sets with variable faces in uncontrolled environment. Sung and Poggio developed a distribution-based approach for FD which was the first accurate appearance-based method [4]. Training examples are gathered from creation of virtual faces and bootstrapping. Each face and non-face is normalized using masking, illumination gradient correction and histogram equalization. All training patterns are grouped into six face and six non-face clusters. Euclidean and normalized Mahalanobis distances are computed between an input image pattern and the prototype clusters. Multilayer perceptron network is applied to classify face window patterns from non-face patterns using the distances to each face and non-face cluster.

The first advanced neural network-based approach that reported results on a large and difficult dataset was by Rowley et al. [5]. It becomes de-factor the standard for evaluation with other upright frontal FD approaches. Their system incorporates face knowledge in a retinally connected neural network, looking at windows of 20x20 pixels. In their single neural network implementation, there are two copies of a hidden layer with 26 units, where 4 units look at 10x10 pixel sub-regions, 16 look at 5x5 sub-regions, and 6 look at 20x5 pixels overlapping horizontal stripes. The input window is pre-processed like in the Sung and Poggio's system [4]. The image is scanned with a moving 20x20 window at every possible position and scale with a subsampling factor of 1.2. To reduce the number of false alarms, they combine multiple neural networks with an arbitration strategy. The fast version of FD system uses extra neural network that scans an image with 30x30 pixels window and 10 pixels step for face candidates which then are passing to the verification neural network.

A new extremely fast FD algorithm is presented by Viola and Jones [3] that uses AdaBoost for selecting essential Haar-like features and the attention cascade of classifiers.

The state of the art methods [3, 5] still have some disadvantages. For example, FD system which is based on [3] misses partially-occluded or hardly shadowed faces and gives more false positive than in [5], whereas FD approach which is described in [5] is too slow for real-time video-flow processing. In our paper we propose to combine the abovementioned approaches to overcome these disadvantages by using some Haar-like features from [3] for face candidate selection and improved FD neural network-based method, adapted from [5]. We also used color segmentation preprocessing stage with image color balance enhancement, skin detection in several color-spaces and morphological operations for the FD process acceleration. After the pre-processing stages the final FD is performed using improved face search strategy across scale and position with the following key elements: inverse image scale pyramid, adaptive window scanning step and window acceptance. These improvements in search strategy allow reducing the number of handled windows especially in the case of large faces presence. Training set for neural network is formed in bootstrap manner not only for non-faces but also for faces. This provides to draw a distinction between two classes more precisely.

The rest of this paper is organized as follows: first, we describe face candidate selection algorithms which are based on skin color segmentation and Haar-like features' analyzing, in section 3 the improved neural network-based method is described in details and in the last section the conclusions and the future directions of our research are given.

2 Face Candidate Selection

2.1 Face Candidate Selection Using Skin Color Segmentation

The human skin has a characteristic color and could be easy recognized by people. Therefore, the usage of skin color (SC) information can considerably facilitate the process of faces exposure, localization and tracking [2]. Color allows fast processing of the input image and is highly robust to geometric variations of the face pattern.

SC segmentation can be based on separate pixels or on regions. In this work we use pixel segmentation, including classifier creation to separate skin-pixels from the background. The classifier creation accomplished by determination of the metrics that measures distances between the pixel color and SC. The metrics type is defined from the SC modeling method: explicitly defined skin region (defining skin region boundaries), nonparametric skin distribution modeling (defining of the skin color distribution from training set), parametric and dynamic skin distribution modeling [2]. We use the method of explicitly defined skin region boundaries as it is simple, fast, and exact enough.

There are a few color spaces which successfully applying for segmentation tasks: RGB, nRGB, HSV, TSL, HSI, YIQ, YCbCr and other. Our experiments show that the best segmentation is provided by the combination of RGB and TSL color-spaces (Fig. 1). We use the follow rule to determine the boundaries of the SC cluster in RGB color space (for each of the R, G, and B channels) [6]:

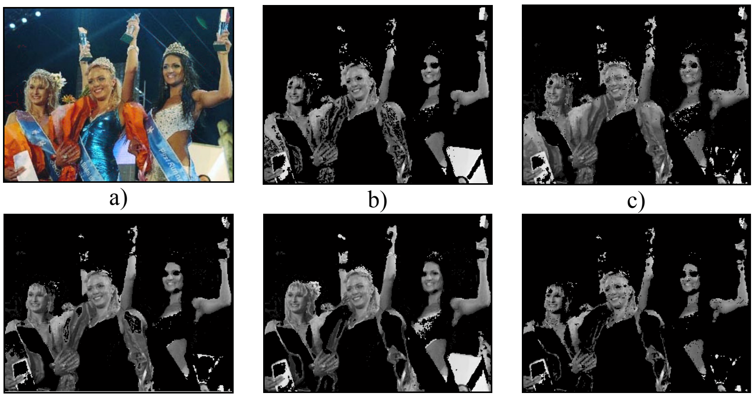
$$\begin{gathered}
\%\ \text{The skin color model at uniform daylight illumination} \\
R > 95 \ \mathit{and}\ G > 40 \ \mathit{and}\ B > 20 \ \mathit{and} \\
\max\{R,G,B\} - \min\{R,G,B\} > 15 \ \mathit{and} \\
|R - G| > 15 \\
\mathit{and}\ R > G \ \mathit{and}\ R > B
\end{gathered},$$

$$OR$$

$$\begin{gathered}
\%\ \text{The skin color model under flashlight lateral illumination} \\
R > 220 \ \mathit{and}\ G > 210 \ \mathit{and}\ B > 170 \ \mathit{and} \\
|R - G| \le 15 \ \mathit{and} \\
R > B \ \mathit{and}\ G > B
\end{gathered}$$

The usage of the additional spaces (YCbCr, YIQ) allow to reject some more background pixels, but the speed of segmentation block executing will fall down.

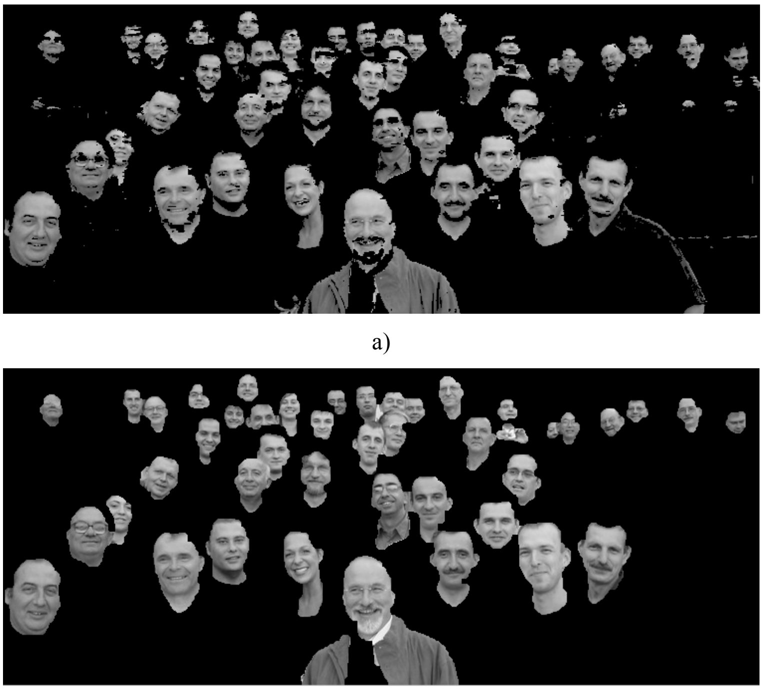
Color balancing is performed before the segmentation to adjust color distribution. The segmentation is followed by the morphological operations (opening, closing, and filtration) in order to improve an image quality (Fig. 2).



a) b) c)

d) e) f)

Fig. 1. Segmentation results of input image (a) using RGB (b), TSL (c), YCbCr (d),YIQ (e) color spaces and the result of their combination (f).



a)

b)

Fig. 2. Segmented image before (a) and after (b) applying of the morphological operations.

SC segmentation allows extremely reduce a face search area and speedup the whole FD process in 5-20 times depending on the input image.

2.2 Face Candidate Selection Using Haar-like Features

We use some Haar-like features, presented in [3], as a preprocessing step to reduce the face search area (Fig. 3). The size and position of these features is selected in order to provide the error less than 1% on the training set. The features also used on the training stage to reduce the number of non-face images, gathered during the bootstrapping.

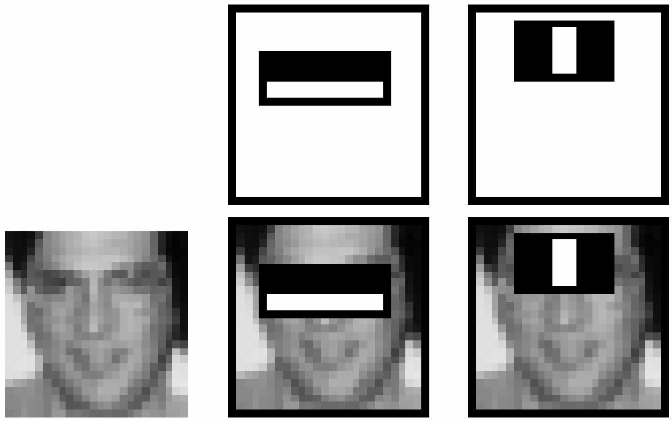


Fig. 3. First two Haar-like features [3].

In comparison with [5] the usage of these features extremely reduces the number of analyzed sub-images for the final classifier (see Section 3).

3 Improved Neural Network-Based Face Detection Method

3.1 Neural Network Active Training Algorithm

The face images for the training set, which were collected from MIT CBCL face data set [7] and Internet, were scaled and cropped to the size of 20x20 pixels. The training set was extended using virtual examples creation by randomly mirroring, rotating, scaling, translating and blurring each of the original face samples. Unlike classical virtual examples creation procedure described in [4, 5] we translate training face samples by 0.5 and 1 pixel vertically and horizontally purposely, to increase the default window scanning step to 2 pixels. We also used blurring operation to extend the training set with cinema-like faces. The total size of the training set is 3242 face images.

We used active training algorithm for retinally connected neural network [5] with a bootstrapping procedure extended on faces where masking, illumination gradient correction and histogram equalization were applying for each of the training sample. Active training algorithm consists of the following steps, adapted from [5]:

1. Create an initial training set by randomly selecting 500 face images from the whole face set and generating 500 random non-face images. Apply the preprocessing steps to each of these images.
2. Train a neural network to produce an output of 0.9 for the face examples and -0.9 for the non-face examples. The training algorithm is a scaled conjugate gradient back-propagation. If mean square error is too large, find the training sample with the biggest error and exclude it from the current training set. Go to step 2.
3. Run the system on images which contain no faces. Randomly collect 25 sub-images in which the network incorrectly identifies faces as negative examples.
4. Run the system on the whole face set. Randomly collect 25 face images in which the network incorrectly identifies non-faces as positive examples. If the number of collected images smaller than 25, randomly select the deficient images from the whole face set.
5. Apply the preprocessing steps to collected face and non-face images and add them to the current training set. Go to step 2.

Such training algorithm provides the network with relatively small representative training set (5440 images after 100 training epochs) since the network is collecting face and non-face examples itself. The testing of the trained neural network was performed on MIT CBCL face test set [7] which includes 472 face and 23573 non-face images and the average error was 1.96%.

3.2 Improved Face Search Strategy Across Pose and Scale

The classical face search strategy (FSS) across pose and scale supposes the gradual decrease of the input image with some scale coefficient and FD is performed by shifting a search window over the input image with some moving step (usually it equals to 1). Then each of the sub-images is classified to face/non-face class using a classifier [4, 5, 8]. We propose to improve the FSS using inverse image scale pyramid, adaptive window scanning step and window acceptance. These improvements allow decreasing the number of sub-images processed by classifier.

The image scale pyramid is constructed from the smallest image (usually equals to scanning window size) to its original size (Fig. 4).

Fig. 4. The image scale pyramid.

First, the neural network-based classifier looks for large faces. When the face candidate region has some number of position and scale detections this face can be accepted and its image region can be eliminated from further processing (Fig. 5). This verification requires the on-line registration of multiple detections during the detection process unlike the off-line detection results processing used in [5].

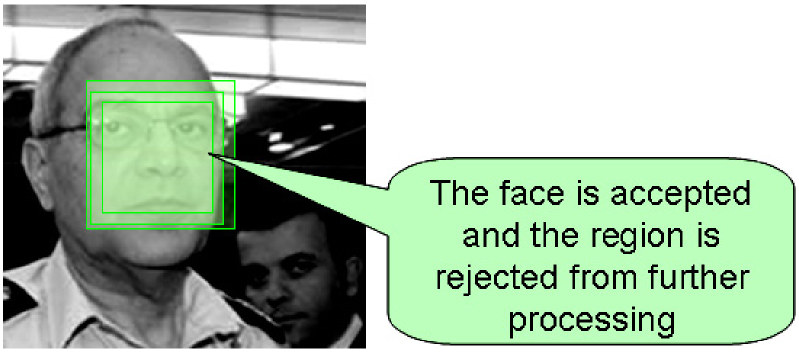


Fig. 5. Face window acceptance.

The classifier avoids analyzing of the accepted face regions using adaptive window scanning step when looking for smaller faces. The default value of adaptive step is 2 (along rows and columns) and it changes in the following cases:

– face-like region (region with a deficient number of multiple detections) is found: the step decreases to 1;

– face candidate is found: the step essentially increases one-time and then sets to its default value;

– accepted region is found: the step essentially increases one-time.

Table 1 shows considerable diminishing of the sub-images number which is analyzed by the neural network using adaptive step and Haar-like features while processing a 71x74 grayscale image (Fig. 5) (experiments are performed in Matlab environment).

Table 1. Face detection using improved face search strategy.

Face search strategy	Number of processed windows
Classical FSS [5]	5792
Improved FSS	295
Improved FSS and 2 Haar-like features	64
Improved FSS and 4 Haar-like features	13
Improved FSS and 6 Haar-like features	7

The improved FSS in conjunction with the application of Haar-like features allows accelerating FD process by diminishing of the scanning sub-images number especially when input images contain large faces.

4 Conclusions and Future Works

This paper presents some face candidate selection algorithms and improved neural network-based method. Face candidate detection is performed using the skin color and Haar-like features. The improved active training algorithm allows neural network working with the relatively small representative training set. The proposed face search strategy accelerates the face detection process using the inverse image scale pyramid, adaptive window scanning step, window acceptance, and is perfectly suitable for input images with large faces.

Our future research will be focused on further speedup of the face search process by construction of classifier's cascade, like in [3], where the final strong classifier (retinally connected neural network) is transformed into the cascade of modular neural networks. We're also transforming our Matlab routines into C++ application using OpenCV library [9].

Acknowledgements

The authors are grateful for the support to the Fundamental Researches State Fund of Ukraine, as the above results were obtained as a part of the research project "Development of methods and algorithms of face detection and recognition for real-time video-supervision systems".

References

1. Ming Hsuan Yang: Recent Advances in Face Detection, IEEE ICPR 2004 Tutorial, Cambridge, United Kingdom (2004)
2. V. Vezhnevets, V. Sazonov, A. Andreeva: A Survey on Pixel-Based Skin Color Detection Techniques, Graphics and Media Laboratory, Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, Russia (2003)
3. P. Viola, M. Jones: Robust Real-Time Face Detection, International Journal of Computer Vision 57(2) (2004) 137–154
4. K. 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) 39-51
5. H. Rowley, S. Baluja, and T. Kanade: Neural network-based face detection. In IEEE Patt. Anal. Mach. Intell., Volume 20 (1998) 22–38
6. Peter Peer, Jure Kovac, Franc Solina: Human Skin Colour Clustering for Face Detection, EUROCON 2003 - International Conference on Computer as a Tool, Eds. B. Zajc, Volume 2. Ljubljana, Slovenia (2003) 144-148
7. CBCL Face Database #1. MIT Center For Biological and Computation Learning. http://www.ai.mit.edu/projects/cbcl
8. Raphael Feraud, Olivier J. Bernier, Jean-Emmanuel Viallet, and Michel Collobert: A Fast and Accurate Face Detector Based on Neural Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 1 (2001)
9. http://sourceforge.net/projects/opencv/