An Automatic Face Detection System for RGB Images

T. Barbu

Tudor Barbu
Institute of Computer Science, Romanian Academy, Iaşi branch, Iaşi, Romania
E-mail: tudbar@iit.tuiasi.ro, www.iit.tuiasi.ro/ tudbar

Abstract: We propose a robust face detection approach that works for digital color images. Our automatic detection method is based on image skin regions, therefore a skin-based segmentation of RGB images is provided first. Then, we decide for each skin region if it represents a human face or not, using a set of candidate criteria, an edge detection process, a correlation based technique and a threshold-based method. A high face detection rate is obtained using the proposed method.

Keywords: color image, color space, RGB, HSV, skin region, face detection, cross-correlation coefficient, edge detection, template matching, threshold.

1 Introduction

This paper approaches an important digital image analysis domain. *Face detection* represents a computer technology that determines the locations and sizes of human faces in arbitrary digital images.

Face detection can be regarded as a specific case of *object-class detection.* In object-class detection, the task is to find the positions and sizes of all objects in an image that belong to a given class [1]. While being a sub-domain of the object detection field, face detection represents a generalization of *face localization.* In face localization, the task is to find the location and size of a given input image, while in face detection one does not have any information about the human faces [2].

The most important application area of face detection is *biometrics.* Face finding is often considered the first step of the *face recognition process* [3, 4]. Thus, most facial recognition systems, and the more complex biometric systems including face recognition components, use face detection techniques. *Video surveillance* represents another important application domain of face detection.

A robust face detection task consists of identifying and locating all the faces in an image, regardless of their position, scale, pose, orientation and illumination [2, 3]. Early face-detection techniques focused on the detection of frontal human faces only, and did not consider the *rotation* problem. The newer methods attempt to solve the more general and difficult problem of multiview face detection. These algorithms take into consideration the two types of face rotation: *pose*, representing the out-of-plane rotation, and *orientation*, representing the in-plane rotation [2].

Also, there are several factors which could transform the human face finding process into a difficult task, such as: the *structural components* (presence or absence of beards, moustaches, glasses or other elements), the *facial expressions* (smiling, laughing, crying and others), *occlusions* (the faces can be occluded by other objects) and *imaging conditions* (lighting, camera characteristics) [2]. A robust face detection approach must take into consideration the presence of these factors.

There are several known categories of face detection approaches: *knowledge-based* techniques [5], *feature-based* methods [2,6,7], *appearance-based* approaches [8–13] *template matching* methods

[14]. The knowledge-based methods encode human knowledge of what constitutes a typical face, usually the relationships between facial features. A face is represented using a set of human-coded rules. These rules are then used to guide the face search process.

The advantages of the knowledge-based techniques are: the easy rules to describe the face features and their relationships, and the good results obtained for face localization in uncluttered background. Their disadvantages are: the difficulty to translate the human knowledge in rules precisely and the difficulty to extend these methods to detect faces in different poses, respectively [5].

The feature based approaches aim to detect invariant face features. These are structural features of a face that exist even when the pose, viewpoint or lighting conditions vary. We could mention here the *Random Graph Matching* based approaches [6] and the *Feature Grouping* techniques [7]. The main advantage of the feature oriented face detection approaches consists in the fact that these features are invariant to rotation changes. Their main drawback is the difficulty to located facial features in a complex background.

Appearance-based techniques train a classifier using various examples of faces. The classifiers which can be used in the training process include: Neural Networks (Multilayer Perceptrons) [8], Hidden Markov Models [9], Bayes classifiers [10], Support Vector Machines (SVM) [11], Sparse Network of Winnows (SNoW) [12], Principal Component Analysis (PCA) [3] and Boosting algorithms (Ada-Boost) [13].

The template matching based techniques use stored face templates [2, 14]. Usually, these approaches use correlation operations to locate faces in images [15]. The templates are handcoded, not learned. Also, these templates have to be created for different poses.

We propose a template matching based face detection method in this paper, too. Our detection technique works for RGB color images only and it is based on the skin regions of the image. Thus, in the first stage, our approach performs a skin segmentation process, extracting the human skin regions from the analyzed image. The proposed skin detection technique is described in the next section.

Next, the technique identifies the human faces, by performing an analysis of the previously obtained skin segments. The face identification method is provided in the third section. In the fourth section, the experiments performed using the proposed human face detection system, are discussed. The paper ends with a conclusions section and the references section.

2   A Skin Detection Approach for RGB Images

Human skin color is proven to represent a very useful face detection and localization tool [2, 14, 16, 17]. A skin-based face finding approach identifies the skin regions of the image, then determine those of them which represent human faces. Besides face detection, there exist other important application areas of skin detection, such as image content filtering and finding illegal internet content [16], content-aware video compression or image color balancing.

Many skin color localization techniques have been developed in recent years. A robust and very known skin finding method is the algorithm proposed by Fleck and Forsyth in 1996, that uses a skin filter [16].

We are interested in color images only and do not perform skin and face detection in grayscale images. Obviously, the color images are usually in the RGB format. While it is one of the most used color spaces for processing and storing of digital image data, RGB is not a favorable choice for skin color analysis, because of the high correlation of its three channels and the mixing of luminance and chrominance data [17]. For this reason, most skin segmentation algorithms work with other color spaces, such as the *normalized* RGB, HSV (and other Hue saturated based spaces) and YC_rC_b formats [2, 17].

We propose a skin detection technique using the HSV and YC_rC_b color spaces. First, a denoising process should be performed on the input RGB image, I. Usually, these images are affected by Gaussian noise. Therefore, a 2-D Gaussian smoothing filter has to be applied to them, to remove detail and noise [18].

Then, the smoothed image is converted into the Hue Saturation Value format, by computing the three components using the known conversion equations. We obtain the components, H, S and V, as three matrices with coefficients in the $[0, 1]$ interval. We are interested mainly in the hue value, H.

The YC_rC_b color model represents a family of color spaces [17]. In fact, it is not an absolute color space, but a way of encoding the RGB information. In this format, Y represents the luminance, while C_r and C_b are the blue-difference and red-difference chroma components. These three components of the color space are computed as linear combinations of R, G and B components of the image.

Thus, the computation formulas of the chroma components, C_r and C_b, have the general form $\alpha \cdot R + \beta \cdot G + \gamma \cdot B + 128$, where coefficients $\alpha, \beta, \gamma \in [-0.5, 0.5]$. We choose empirically some proper values for these coefficients and get the following components:

$$\begin{cases} C_r & = & 0.15 \cdot R - 0.3 \cdot G + 0.45 \cdot B + 128 \\ C_b & = & 0.45 \cdot R - 0.35 \cdot G - 0.07 \cdot B + 128 \end{cases} \tag{1}$$

Then, we apply a set of restrictions on these two components and on the hue, to identify the skin regions. Thus, we have determined a skin related interval for each component. In our approach, each pixel of the image I belongs to a human skin segment if the corresponding values in C_r, C_b and H are situated in those intervals.

We create a binary image Sk, having the same size as I, whose white regions correspond to the skin segments. The proposed skin segmentation process is modeled by the following relation:

$$Sk = \begin{cases} 1, & if\ C_r(i, ju) \in [150, 165] \propto C_b(i, j) \in [145, 190] \propto H(i, ju) \in [0.02, 0.1] \\ 0, & otherwise \end{cases} \tag{2}$$

where $i, j \in [1, M]$ and $j \in [1, N]$, I representing an $[M \times N]$ image.

The connected components from image Sk, computed by (2), represent the detected skin regions. The proposed detection method provides good results, although some skin identification errors could appear. That means some non-skin image regions could be detected as skin segments, but this fact will not affect the final goal, human face detection. Therefore, we are satisfied with the obtained skin finding results.

In Figure 1 (a), there is displayed an RGB image depicting human persons. The result of the HSV conversion is displayed in Figure 1 (b).

The skin detection process is performed by applying the equations (1) and (2). The resulted skin regions are those depicted in Figure 2.

3   A Face Finding Technique

The skin regions detected in the previous section are used in the human face identification process. For each skin segment we have to decide if it represents a face, or it is a non-facial skin region. We will propose an automatic template matching scheme for face detection. Before applying the matching procedure, our face finding approach performs several necessary pre-processing steps.
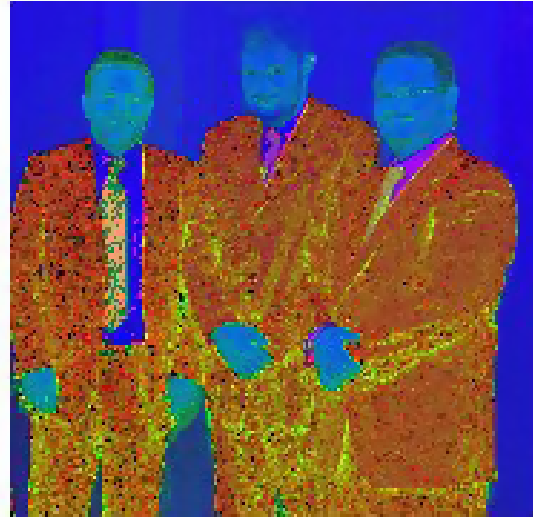
Figure 1: Digital color image conversion: RGB to HSV

Figure 2: Skin detection result

3.1 Skin region pre-processing

A task we try to solve is the separation of human faces from adiacent or occluding skin regions. Our detection task cannot identify properly the faces which are occluded by other skinlike *objects* in the images. Usually, this situation appears in group photos, like that displayed in Figure 3 (a).

Therefore, we provide a separation technique involving some morphological operations [19], performed on the corresponding binary image, Sk. Thus, we will apply two successive *erosions* on it. First, the binary image is eroded with a structuring element L, representing a vertical line, having a length of 5 pixels:

$$Sk' = Sk \ominus L = \bigcap_{\ell \in L} Sk_{-\ell} \tag{3}$$

Figure 3: Face Separation

Then, another erosion operation is performed on Sk', using a structuring element Sq, representing a small square area (for example containing a single pixel):

$$Sk'' = Sk' \ominus Sq = \bigcap_{p \in Sq} Sl'_{-p} \tag{4}$$

In the figure above, one can see a skin separation example using the proposed method. In Figure 3 (b), the skin segments corresponding to the faces from Figure 3 (a) are conjoined into a single region. The result of the morphology-based process, given by the relations (3) and (4), is displayed on Figure 3 (c). The two greatest skin regions are clearly separated in the final binary image, Sk''. Obviously, there could be situations, provided by big occlusions for example, when the face separation is not possible in the binary image.

The binary image Sk'' contains a set of skin segments, which represent connected sequences of white pixels. Let this set of regions be $\{S_1, ..., S_n\}$. Now, we have to decide which of these regions could qualify as *face candidates* for the template matching process. So, we have established a set of candidate criteria for S_i segments.

First condition is size related. The skin region set is usually very large because of the many small white regions which could be present in image Sk''. We have decided to not take into consideration these small area white spots, because they cannot represent serious face candidates.

It is still possible to exist very small faces in an image, like those in a very large crowd, but we consider them irrelevant and do not try to detect them. So, if a white region area (the number of pixels) is below a given threshold value, that region is labeled as a non-facial segment.

Another condition is related to the shape of the skin regions. Obviously, a face region should have a rectangular-like or ellipse-like shape. Thus, a rigorous shape analysis can determine which skin segments cannot represent human faces.

We propose a less complex approach to this task. A connected component S_i has to be rejected as non-face, having a non-facial shape, if its *solidity*, representing the ratio between its area and its *bounding box* area is below an established threshold. A facial skin-region is usually characterized by a high ratio, close to 1, but the solidity value may become lower because the region's area is affected by the presence of many black holes in the face region. These holes could represent human face components, such as eyes, eyebrows, mouth, nose, ears and wrinkles, or some skin detection errors. Therefore, we perform a black hole filling process on the binary image Sk'', first, then compute and use the areas of the filled S_i segments.

Also, a human face is characterized by some limits of its *width to height ratio*. None of the two dimensions of a face, the width and the height, can be *much* larger than the other. For this reason, we set another condition, requiring the width to height ratio of the face candidates to be restricted to a certain interval. The ratio between the width of the region's bounding box and the height of the bounding box should be between two properly chosen threshold values.

The white regions satisfying the proposed restrictions represent the face candidates. For each S_i, $i \in [1, n]$, the described face candidate identification process is formally expressed as follows:

$$Area(S_i) \geq T_i \alpha \frac{Area(Fill(S_i))}{Area(Box(S_i))} \geq T_2 \alpha \frac{Width(Box(S_i))}{Height(Box(S_i))} \in [T_3, T_4] \Longrightarrow S_i = candidate \quad (5)$$

where $Area(\ )$ computes the number of white pixels of the region received as argument, $Fill(\ )$ performs the filling process, $Box(\ )$ returns the bounding rectangle, $Width(\ )$ and $Height(\ )$ returning the dimensions of a rectangle. We have considered the following proper values for the thresholds in equation (5): the area threshold $T_1 = 130$, the solidity threshold $T_2 = 0.65$, and the width to height thresholds $T_3 = 0.6$ and $T_4 = 1.8$.

A face candidate identification example is described in Figure 4. In the RGB image from Figure 4 (a) one can see a boy flexing his muscles. Figure 4 (b) represents the corresponding binary image resulted after skin detection, erosion operations, small region removing and hole filling process. The bounding boxes of the three remaining skin regions are depicted in pictures c), d) and e). The one representing the right arm is rejected because of its low solidity percent (meaning a wrong shape), the skin segment of the left arm is rejected because a wrong width to height ratio, and the one representing the skin of head and neck is accepted as a right face candidate.

3.2   Template matching process

In its next stage, our facial detection approach determines which of the face candidates represent human faces. First, one converts the denoised RGB image I into a $2D$ grayscale form, let it be I'. If there is a set of K face candidates, where $K \leq n$, then we determine the set of the sub-images of I' corresponding to the bounding boxes of these candidates.

The face detection process can be affected by the head hairline of the persons and by the skin zone corresponding to the neck and the upper chest. Therefore, a narrow upper zone and a narrow bottom zone from each image are removed. In our tests, the height of each removed zone represents one eleventh of the bounding box height.

Let the set of the truncated *skin images* be $\{I_1, ..., I_K\}$, where $I_i \subset I'$, $\forall i \leq K$. Then, we perform a correlation-based template matching process on this set. Our template-based approach
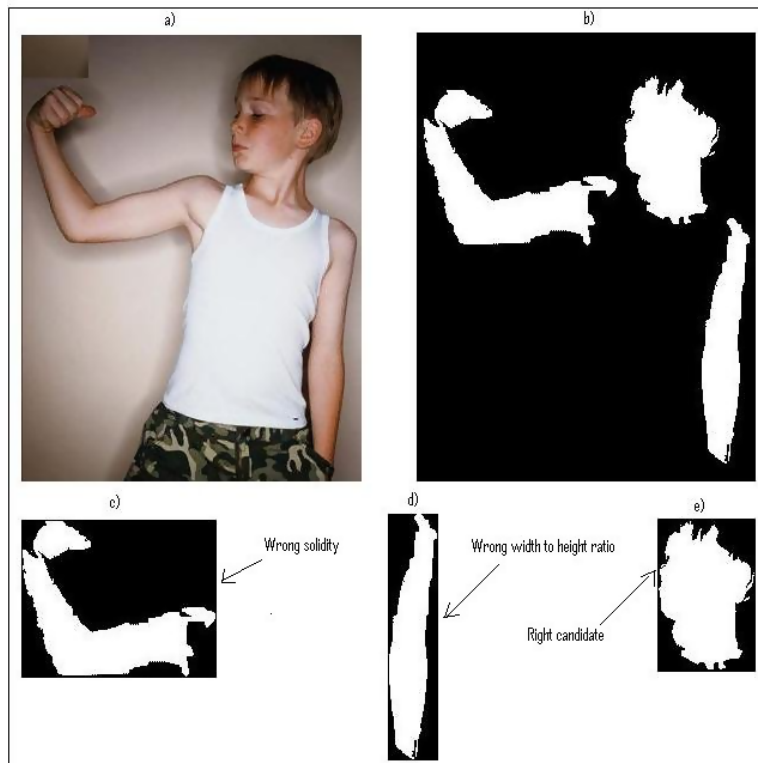
Figure 4: Face candidate identification example

works like a supervised classification algorithm. We create a face template set, containing human faces of various sizes, orientations and poses, and representing both male and female people, of various ages and races. Let the template set be $\{F_1, ..., F_N\}$, with N large enough, where each F_i represents a grayscale image.

Next, an edge detection operation is performed on both the skin image set and the face template set. A Canny filtering technique is used for image edge extraction, because this detector is less likely than the others to be affected by noise [20]. First, it computes the gradient of the image, using the derivative of a Gaussian filter. Then, it finds edges by looking for local maxima of this gradient. The Canny method uses two thresholds, to detect strong and weak edges. It takes into consideration only the truly weak edges, representing those connected to the strong ones. Thus, for each skin image I_i and each face image F_j, a binary image representing its edges is determined. Let us note I_i^e and F_j^e the *edge images* corresponding to images I_i and F_j respectively.

Then, for each candidate (skin image), one computes the $2D$ *cross-correlation coefficients* [21] between its edge image and the edge images of the templates, and the average value of this sequence of coefficients. Let us note $v(I_i)$, this mean value corresponding to image I_i. Each time a correlation operation is performed the edge image of the candidate has to be resized to the size of the template. The best solution to the face detection task is a threshold-based one. The computed two-dimensional mean correlation coefficient corresponding to a facial skin image, must exceed a properly chosen threshold value. The face identification process is expressed mathematically as follows:

$$\forall i \in [1, K], \quad I_i = face \Longleftrightarrow v(I_i) \geq T \tag{6}$$

where T represents the chosen threshold and

$$v(I_i) = \frac{1}{N} \sum_{j=1}^{N} \frac{\Sigma_x \Sigma_y (I_i^e(x,y) - \mu(I_i^e))(F_j^e(x,y) - \mu(F_j^e))}{(\Sigma_x \Sigma_y (I_i^e(x,y) - \mu(I_i^e))^2)(\Sigma_x \Sigma_y (F_j^e(x,y) - \mu(F_j^e))^2)} \tag{7}$$

where $\mu(\ )$ computes the mean of a matrix. The threshold value is determined empirically. From the performed experiments, we have got a satisfactory threshold value, $T = 0.185$. If T is not exceeded for any I_i, then the color image I contains no human faces. We propose a no-threshold automatic face finding approach, too. The threshold can be replaced with a clustering procedure that uses the values $v(I_i)$ computed by (7) as feature vectors. Thus, the value set $\{v(I_1), ..., v(I_k)\}$ is divided into two classes. A region-growing algorithm can be used in this case.

There is also a more simple way to perform the clustering: to sort the set in ascending order, then to find the greatest difference between two successive values. That pair of successive correlation-based values marks the dividing point between the two clusters. Obviously, the cluster containing the high values is the one corresponding to the facial images. This method works satisfactory when the values $v(I_i)$ related to faces are *much* greater than those corresponding to non-facial images. Some face detection errors could be produced. For this reason, the threshold based approach is preferred by us.

After determining those I_i images representing faces, if such images exist, the corresponding facial sub-images of the RGB image I are provided to the output of our face detection system. Let us return to the example described in the first two figures. If the described face finding technique is applied to the skin detection results depicted in Figure 2, we get the results displayed in Figure 5.

In Figure 5 (a) there are displayed the main skin regions of the image, obtained after performing the morphological operations, the hole filling and the small region removing processes on the binary image depicted in Figure 2. One of these skin regions is rejected because of its low solidity, the remaining regions being accepted as face candidates, as one can see in Figure 5 (b).

The template matching process is performed using the template face set represented in Figure 6. One can see the resulted average correlation coefficient values in Figure 5 (c), those greater than 0.185 corresponding to the detected faces, surrounded by black rectangles in that grayscale image. The final face detection result for the RGB image is displayed in Figure 5 (d), the human faces being marked by red bounding boxes.

4 Experiments

We performed a lot of face detection experiments using the described system. Our tests involved tens of RGB images containing human faces and produced satisfactory results. A high face detection rate is obtained.

We created a template face set that contains 25 grayscale images of various scales and imaging conditions, for our experiments. As one can see in the pictures below, these templates represent both male and female faces, and people of various ages and races. These faces are also characterized by various orientations and poses, some of them have structural elements, too. The template set can be extended, by adding new faces, but although a large set could improve the detection results, it also produces a high computation complexity.

As mentioned in the previous section, the threshold-based face finding approach provides much better results than the clustering-based one. Our face detection technique is characterized not only by a high detection rate, that is approximately 90% and indicates a low number of *false negatives* (missed faces), but also by a low number of *false positive* (non-facial image regions declared to be faces).
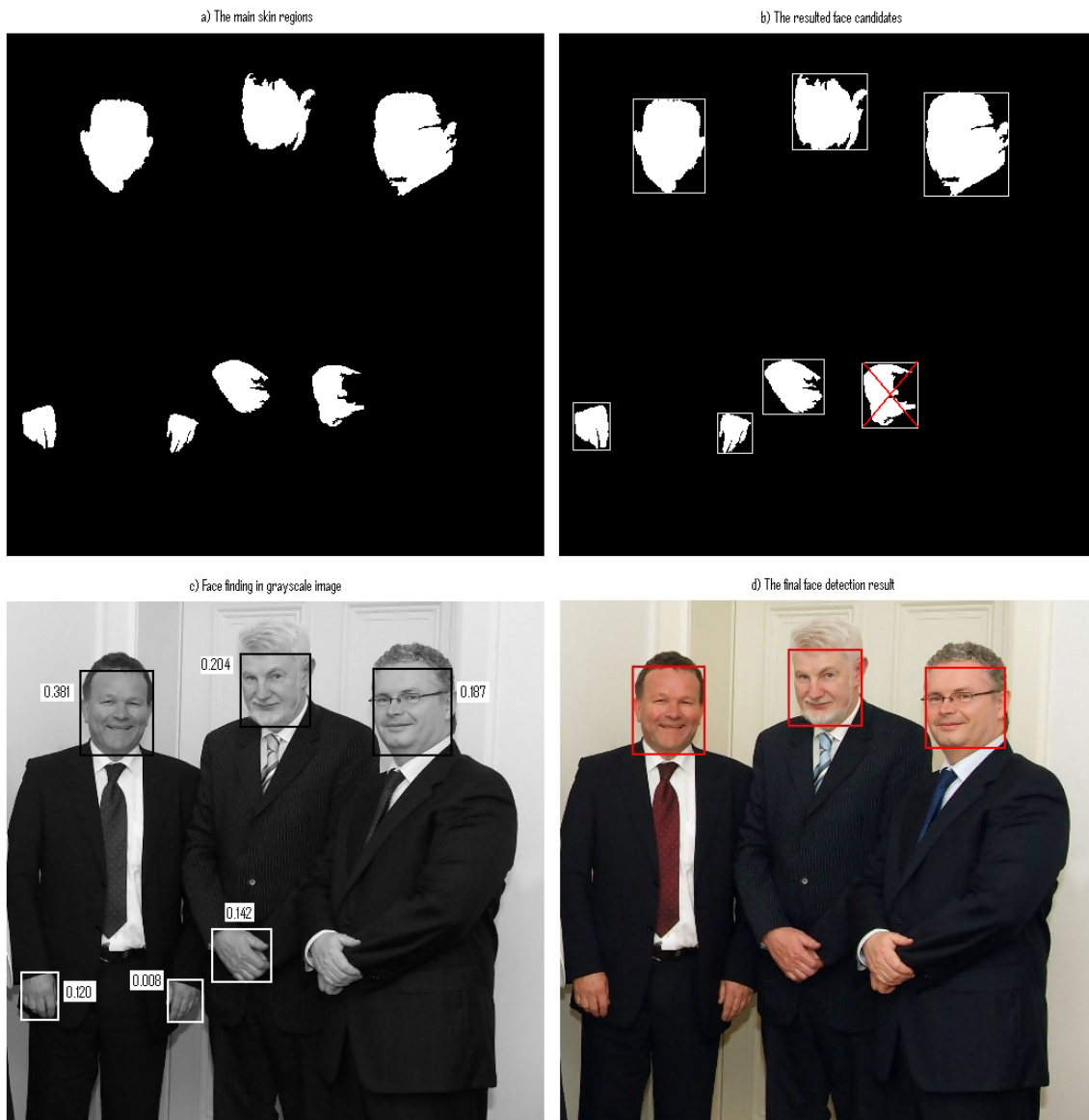
Figure 5: Face detection example

The performances of the proposed face detection system are comparable with those of the face detection approaches mentioned in the introduction. It achieves better detection results for the frontal faces, than for faces characterized by various orientations.

5   Conclusions

A skin segmentation based face detection system for RGB color images has been proposed in this paper. The main contributions of this work are the proposed skin detection approach and face identification technique.

The skin regions resulted from the skin segmentation process are analyzed to determine which of them could represent human faces. A set of face candidate criteria was proposed by us, to reduce the set of face candidates and the computation volume and complexity.

We used a template matching method for face detection and provided a cross-correlation

Figure 6: Template face set

based skin region discrimination approach. Unlike other template matching algorithms, our procedure uses the edges of the skin region and not the region itself. We choose to perform an edge detection first, because we consider that face features are contained mainly in the edge image.

The good face detection results obtained in our experiments prove the effectiveness of our technique. Our future work will focus on developing robust face recognition and more complex biometric systems, using the face detector proposed in this paper. We approached the face recognition domain in our previous works [4], and we want to unify the research in these two areas.

Acknowledgment

Bibliography

[1] C. Papageorgiou, M. Oren, T. Poggio. A General Framework for Object Detection, *International Conference on Computer Vision*, Bombay, India, pp. 555-562, Jan. 1998.

[2] M.H. Yang, D. Kriegman, N. Ahuja. Detecting Faces in Images: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (*PAMI*), Vol. 24, no. 1, pp. 34-58, Jan. 2002.

[3] S. Atsushi, I. Hitoshi, S. Tetsuaki, H. Toshinori. Advances in face detection and recognition technologies, *NEC Journal of Advanced Technology*, Vol. 2, no. 1, pp. 28-34, 2005.

[4] T. Barbu. Eigenimage-based face recognition approach using gradient covariance, *Numerical Functional Analysis and Optimization*, Volume 28, pp. 591 . 601, Issue 5 & 6, May 2007.

[5] G. Yang, T.S. Huang. Human face detection in a complex background. *Pattern Recognition*, Vol. 27, no. 1, pp. 53-63, 1994.

[6] T.K. Leung, M.C. Burl, P. Perona. Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching, *Proceedings of the 5th International Conference on Computer Vision*, pp. 637-644, Cambridge, Mass., June 1995.

[7] K.C. Yow, R. Cipolla. A probabilistic framework for perceptual grouping of features for human face detection, *Second IEEE International Conference on Automatic Face and Gesture Recognition* (FG '96), pp. 16, 1996.

[8] H.A. Rowley, S. Baluja, T. Kanade. Neural Network-Based Face Detection, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 203-208, 1996.

[9] A.V. Nefian. An embedded HMM-based approach for face detection and recognition, *Proceedings of the Acoustics, Speech, and Signal Processing ďž˝e99 on 1999 IEEE International Conference*, Vol. 6, pp. 3553-3556, 1999.

[10] T.V. Pham, M. Worring, A.W.M. Smeulders. Face Detection by Aggregated Bayesian Network Classifiers, *Machine Learning and Data Mining in Pattern Recognition*, Book Series Lecture Notes in Computer Science, Volume 2123, pp. 249-262, 2001.

[11] E. Osuna, R. Freund, F. Girosi. An improved training algorithm for support vector machines, In *Proceedings of IEEE NNSP'97*, pp. 276-285, Amelia Island, Florida, 1997 (a).

[12] M. Nilsson, J. Nordberg, I. Claesson. Face Detection using Local SMQT Features and Split Up SNoW Classifier, *IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP*), Vol. 2, pp. 589-592, April 2007.

[13] K. Ichikawa, T. Mita, O. Hori. Component-based robust face detection using AdaBoost and decision tree, *Proc. of the 7th Int. Conference on Automatic Face and Gesture Recognition*, pp. 413-420, 2006.

[14] Z. Jin, Z. Lou, J. Yang, Q. Sun. Face detection using template matching and skin-color information, *Advanced Neurocomputing Theory and Methodology*, Vol. 70, Issues 4-6, pp. 794-800, Jan. 2007.

[15] S. Majed, H. Arof. Pattern correlation approach towards face detection system framework, *Information Technology, 2008. ITSim 2008. International Symposium on*, Vol. 4, pp. 1-5, Aug. 2008.

[16] D. A. Forsyth, M. M. Fleck. Identifying nude pictures, *IEEE Workshop on the Applications of Computer Vision '96*, pp. 103-108, 1996.

[17] V. Vezhnevets, V. Sazonov, A. Andreeva. A Survey on Pixel-Based Skin Color Detection Techniques, In *Proceedings of the GraphiCon 2003*, pp. 85-92, 2003.

[18] L.G. Shapiro, G. C. Stockman. *Computer Vision*, pp. 137- 150, Prentince Hall, 2001.

[19] H.J.A.M. Heijmans. Morphological Image Operators, *Advances in Electronics and Electron Physics*, Boston: Academic Press, 1994.

[20] J. Canny, A Computational Approach To Edge Detection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 88, pp. 679-714, 1986.

[21] A.L. Edwards, *An Introduction to Linear Regression and Correlation*, San Francisco, CA: W.H. Freeman, pp. 33-46, 1976.