Face Detection in Still Color Images Using Skin Color Information

Aouatif Amine, Sanaa Ghouzali and Mohammed Rziza

***Abstract*— In this paper, we develop two face detection methods in color images. These methods rely on a two step process. First, we detect human skin regions using Bayes rule in color images. To avoid the effect of brightness included in the RGB color space, we propose to model the skin color in the chromatic and pure color space YCrCb, which separates luminance and chrominance components. A Gaussian probability density is estimated from skin samples, collected from different ethnic groups, via the maximum-likelihood criterion. To localize the faces within the detected skin regions, two different face detection techniques are then performed. The first algorithm is based on a Template Matching method using the Support Vector Machines "SVM". The second one consists on a multi-layer Neural Network classification. In the following, experiments are carried out and satisfactory results are obtained which indicate the robustness of the first process to detect faces under different environment conditions.**

I. INTRODUCTION

Face detection is a well-known pattern recognition problem. Such task is the first fundamental step for many applications such as face recognition and 3D face reconstruction. Although many approaches have been proposed over the last few years, it still remains a very challenging problem today [1][2] due to significant face appearance variations, such as pose (front, non-front), occlusion, image orientation, lighting conditions and facial expression.

Many researches have been done in this area, including that developed by Rowley and al [3] who propose a neural network system able to detect faces of different sizes and rotations at each image position. Osuna [4] uses support vector machine with polynomial kernel to build a face/non-face classifier by maximizing the margin between the two separated classes. Leung and al [5] use a "graph matching" method to find probable faces from detected facial features. These graphs are generated from the obtained features and the true faces are detected among the candidates by random graph matching. Recently, Viola and al [6] propose a face detection method that computes speedily features using an integral image and combine classifiers in a cascade allowing rejecting background regions quickly. Fleuret and al [7] present a coarse-to-fine face detector which is entirely based on edge configurations. This algorithm visits a hierarchical partition of the face pose space, and in order to declare one detection, a chain of classifiers from the root to one leaf is found. Their learning approach is purely statistical.

In this paper we propose two automatic methods for frontal face detection in still color images. These methods are based on a two image processing steps which first detect regions which are likely to contain human skin in the color image and then extract information from these regions which might indicate the location of a face in the image. The skin detection is performed using a Bayesian framework. Faces are then detected on image containing only the extracted skin areas using two different algorithms. The first algorithm consists on a template matching method based on the SVM technique where the template is the average of the face support vectors. A correlation between the determined template and the candidate skin regions is calculated to indicate the presence of the face(s). The second face detection algorithm is constructed based on a multi-layer Neural Network with back-propagation.

The remainder of this paper is organized as follows: The skin detection algorithm is described in section 2. In section 3 we present two face detection systems based on template matching and the neural network classification, respectively. Then we discuss these methods in section 4. Finally in section 5, a conclusion is given.

II. SKIN DETECTION

Human skin color has been used and proven to be an effective feature for face detection, localization and tracking. Although different people have different skin color, several studies have shown that the major difference lies largely between their intensity rather than their chrominance [8][9]. Many techniques [10][3] have been reported for locating skin color regions in the input image. While the input color image is typically in the *RGB* format, these techniques usually use color components in other color spaces, such as the *HSV* or *YCrCb* formats. That is because RGB components are subject to the lighting conditions which may induce face detection to fail. Among many color spaces, we used *YCrCb* components since the luminance information is contained in Y component, and the chrominance information is in Cr and Cb. Therefore, the luminance information can be easily de-embedded.

In the skin color detection process, each pixel was classified as skin or non-skin based on its color components where skin color distribution is modelled by a Gaussian joint probability density function, defined as:

$$P(Y/w_s) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma_s)^{\frac{1}{2}}} \exp(-\frac{1}{2}(Y-\vec{\mu}_s)^\top \Sigma_s^{-1}(Y-\vec{\mu}_s)) \quad (1)$$

Here, Y is a color vector and μ_s and Σ_s are the distribution parameters (mean vector and the covariance matrix of skin class w_s, respectively), d is the dimension of the color vector.

A. Amine, S. Ghouzali and M. Rziza are with Groupe Signaux, Communications & Multimédia (GSCM), Faculté des Sciences, B.P. 1014, Rabat, Morocco. {rziza}@fsr.ac.ma

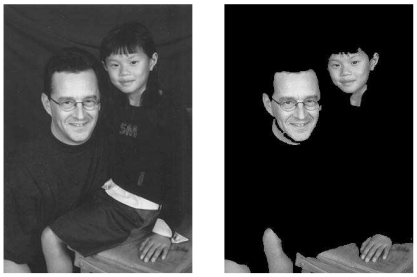


Fig. 1. Skin detection process.

The model parameters are estimated from 18 training faces from different ethnic groups by:

$$\mu_s = \frac{1}{n}\sum_{j=1}^{n} Y_j; \qquad \sigma_s = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j-\mu_j)(Y_j-\mu_j)^{\top}$$

where n is the total number of skin samples Y_j in the training set.

Through the application of Bayes decision rule, the proposed system decides whether a specified pixel Y in the inspected image belongs to the skin class using its $P(Y/w_s)$ probability.

$$Y \in w_s \qquad if \qquad P(Y/w_s) \geq Threshold \tag{2}$$

where "Threshold" is selected empirically from the training of samples.

The skin detection system has been successfully applied and its results are shown in Fig. 1. Some non-skin objects are inevitably observed in the result as their color falls into the skin color space. Indeed, a skin region doesn't always represent a face, and therefore candidate area should be further normalized and checked in order to discern whether it represent a face or not.

III. FACE DETECTION

The face detection involves taking the detected skin areas. To localize the face within a skin region, we implement two different methods. The former is based on template matching combined with the support vector machine technique. One of the most important characteristics of this method is that it uses a human face template to make the final decision of whether a skin region represents a face. The latter is based on Neural Network classification that relies on learning algorithm to find the relevant characteristics of face and non-face images. This method is straightforward because training the neural network uses a huge set of images which should capture the representative variability of facial appearance.

A. Template Matching

To detect faces from the skin regions, the proposed system proceeds into two steps.

In the first step, we determine the candidate skin regions that should be passed to final detection by performing several basic image analysis techniques and using the following hypothesis:

Fig. 2. Human face Template

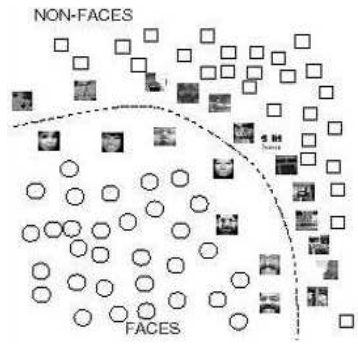


Fig. 3. Support Vector representation[4]

1) a skin region should have at least one hole inside it, therefore, we eliminate those regions that have no holes inside.
2) a height to width ratio must be respected by selected regions to represent a human face.

In the second step, we construct a human face template, as shown in Fig. 2, by averaging 36 "Vectors" that represent frontal view faces of males and females, resulting from the SVM technique [4] on the MIT[1] training data set which is available over the World Wild Web [13]. An interesting characteristic of the SVs is that they lie geometrically nearby the decision plane, as it is seen in Fig. 3. Using these SVs that contained different information of all frontal faces in the training set images, allows us to determine a template which is in effect a 'general' face that could be anybody's face and has most of the features of a face without being too specific.

Then we compute the cross-correlation value between the part of the image corresponding to the skin region and the template face, a way to determine this value is to use the following equation:

$$corr = \frac{\sum_n \sum_m (A_{nm}-\bar{A})(B_{nm}-\bar{B})}{\sqrt{\sum_n \sum_m (A_{nm}-\bar{A})^2 (B_{nm}-\bar{B})^2}} \tag{3}$$

$$\bar{A} = \frac{1}{MN}\sum_n \sum_m A_{nm} \tag{4}$$

$$\bar{B} = \frac{1}{MN}\sum_n \sum_m B_{nm} \tag{5}$$

Where A_{nm} and B_{nm} are respectively matrices of a human face template and detected skin region.

From our experiments, an optimal threshold value of 0.6 was found to work quite well. Therefore, to classify a region as a face, the resulting correlation value must be greater than 0.6. The final product of the original color image is displayed

[1]The MIT CBCL Face Database

with rectangles showing each of the detected faces in the image as represented in Fig. 5.

B. Neural Network

To distinguish head from other objects that have color values similar to skin color, and to identify which of the skin regions detected from the bayesian system represent faces, we implement a neural network with one hidden layer. The proposed neural network models the appearance of upright, frontal views of faces in a scene based on intensity and extracts object features that would indicate the presence of a face [3]. The neural network receives as input a 19×19 pixel window of a skin region, and generates an output ranging from 1 to -1, signifying the presence or the absence of a face, respectively. To detect faces anywhere in the input, the neural network is applied at every location in the skin region. To detect faces larger than the window size, the input image is repeatedly reduced in size, and the neural network is applied at each size (Multi-scale detection).

To compensate for certain sources of image variation, a preprocessing step, adapted from [11], is applied to a window of the skin region. The window is then passed through a neural network, which decides whether the window contains a face. The preprocessing step attempts to equalize the intensity values in across the window. The image-processing operation of our system consists of three distinct parts.

- *Masking:* reduces the unnecessary background noise in a face pattern.
- *Background subtraction:* finds the best fit brightness plane and then subtracted from it to reduce heavy shadows caused by extreme lighting angles.
- *Histogram equalization:* is performed to expand the range of intensities in the window.

Supervised training of a neural network for the face detection task is quite challenging and requires a large number of face and non-face images. However, it's much harder to get a "representative" set of non-face images. The size of the training set for this class can grow very quickly. In this case, the MIT image database which contains (2429 faces and 4548 non-faces) were used as a training set. The training algorithm is standard error back-propagation with momentum [12]. Training via Back-propagation involved determining the sensitivity of the network output to each weight in the network and adjusting each weight in accordance with how much error it caused in the output.

IV. EXPERIMENTAL RESULTS

The neural network-based face detection system has been applied to several test images, and satisfactory results have been obtained. The test set consists of a total of 24025 face and non-face images given in the MIT database. The system shows 93.28% of right hit rate. Fig. 4 represents an output image of our system, which was not used during the training phase. The key

disadvantages are that they train slowly and require a large amount of training data.

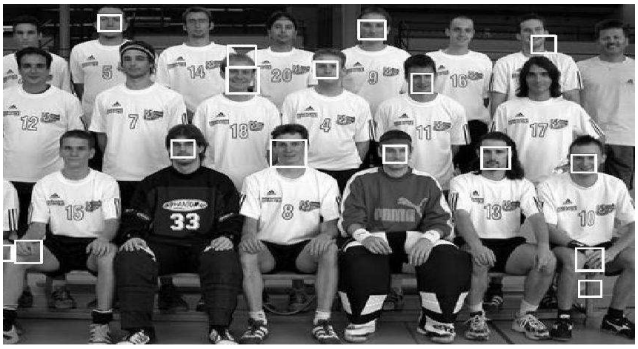


Fig. 4. face detection using Neural Network

To reduce significantly the computational complexity, and improve the overall detection accuracy of the system, more sophisticated features need to be added in order to use it for more general applications. For example, adaptive shape analysis, as the eye/mouth localization approach, can be used to separate isolated human faces from initial segmentation results.

The template matching-based face detection effectively finds instances of front faces in color images. The system works very well with images with only one person and a solid background. But it drags when the background color is similar to skin color. Most of the miss-detections included regions that had very similar skin likelihood values and regions that were indeed skin regions (such as the arms and legs) with more than one hole in the upper part of it.

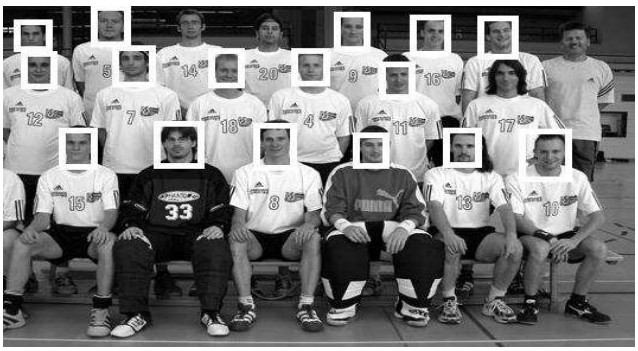
This problem also occurs in connected regions that contain more than one face. In order to separate regions corresponding to isolated faces, a series of morphological operations would have to be used to reliably detect faces in an image. The experimented results are showing in Fig 5.

Template matching-based face detection is simple to implement and detects large, small and half profile faces. We performed extensive experiments to check the validity of the algorithm for subjects with the structural components such glasses, as it seen in Fig 5(b). Whereas the neural network-based face detection system is very complicated and needs to train lots of positive and negative examples to accurately generate the decision function. Moreover, if given a large input image, detecting faces must be done over space and scale, which makes this algorithm very time consuming and impractical for real-time applications.

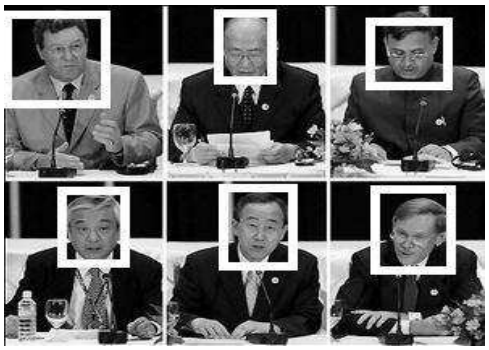
V. CONCLUSION

We have presented two approaches to automatic detection of human faces in color images. The proposed approaches consist of two parts: a human skin segmentation to identify probable regions corresponding to human faces; and a view-based face detection to further identify the location of each human face. The human skin segmentation employs a model-based approach to represent and differentiate the background colors and skin colors. The face detection method based on

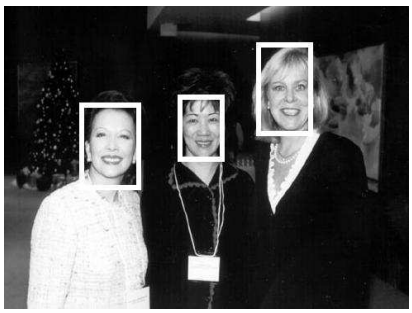
template matching and *SVM* classification is a promising approach given its satisfactory results. The algorithm retains its high performance even in the presence of structural objects like beard, spectacles, mustaches, glasses etc. However, more features with sophisticated algorithm need to be added in order to use it for more general applications.



(a) face detection in image with a complex background



(b) face detection in faces of structural objects



(c) face detection in faces of same scales

Fig. 5. (a),(b) and (c) the results of face detection using Template Matching-SVM

REFERENCES

[1] M-H Yang, D. Kriegman and N. Ahuja, "Detecting Face in Images: A Survey", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, January, 2002, pp. 34-58.

[2] E. Hjelms and B-K Low, "Face Detection: A Survey", *Computer Vision and Image Understanding*, vol. 83, 2001, pp. 236-274.

[3] H. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection", *In IEEE Trans on PAMI*, vol. 20, no. 1, 1998, pp. 23-38.

[4] E. Osuna, R. Freund and F. Girosi,"Training support vector machines: an application to face detection", *IEEE CVPR*, 1997, pp. 130-136.

[5] T. Leung, M.C. Burl and P Perona, "Finding faces in cluttered scenes using random labelled graph matching", *ICCV*, 1995.

[6] P. Viola and M. Jones, "Robust real-time object detection", *Second International Workshop On Statistical and Computational Theories of Vision-Modeling, Learning, Computing and Sampling*, 2001.

[7] F. Fleuret and D. Geman, " Coarse-to-fine visual selection", *IJCV*, vol. 41, no. 2, 2001.

[8] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating Faces and Facial Parts", *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, 1995, pp. 41-46.

[9] H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan, "Multimodal System for Locating Heads and Faces", *Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, 1996, pp. 88-93.

[10] H. Martin Hunke, "Locating and tracking of human faces with neural networks", *Master's thesis, University of Karlsruhe*, 1994.

[11] John Hertz, Anders Krogh, and Richard G.Palmer, "Introduction to the Theory of Neural Computation", *Addision-Wesley Publishing Company, Reading,Massachusetts*, 1991.

[12] Tom M. Mitchell, "Machine Learning", *McGraw-Hill*, isbn. 0070428077, March 1997.

[13] http://www.ai.mit.edu/projects/cbcl/software-datasets/FaceData2.html.