

# Face detection in image with complex background

Jiri Prinosil<sup>1</sup>, Jan Vlach<sup>1</sup>

<sup>1</sup> Brno University of Technology, Faculty Electrical Engineering and Communication,  
Department of Telecommunications,  
Purkynova 118, 612 00 Brno, Czech Republic  
[xprino01@stud.feec.vutbr.cz](mailto:xprino01@stud.feec.vutbr.cz), [jan.vlach@phd.feec.vutbr.cz](mailto:jan.vlach@phd.feec.vutbr.cz)

**Abstract.** This paper deals with an approach to detect face from image with complex background. That can be use in many areas (eg. security systems, biometrics, telecommunications, etc.).In the first part there is presented brief overview of common used method for face localization in image. On the basis of comparison of its properties there was developed appropriate combination of several methods with particular improvements. The goal was to find suitable accuracy to speed rate. Methods were implemented in MATLAB and accuracy rate was tested on Georgia Tech face database.

## 1 Introduction

In recent years the price of computer and forcing-off techniques decreases, on the other side the quality improves, which faces to its massive expansion. Computers must be able to interact with surrounding environment very often, especially with a man. These reasons result in increasing interest in the field of pattern and object recognition with the use of computers. Perhaps, the most energy was devoted to human faces recognition. Mainly for the reason this interest is useful to many spheres (biometrics, sight systems, user interface, content image coding), also for the complexity of all problems with many ambiguous solving. In telecommunication it can be used in video-calls, where the scanning device positioning is controlled on the base of a head motion tracking. Other application can be in video-calls too, where only the face is transmitted and the own background is added by user. There are more other fields in communication where the face localization can be used (java applications in mobile phones, security systems in wireless communication), All of this lead to more comfortable and more attractive use of wireless networks.

Face detection is first step in systems for face recognition and its pertinent analysis. All next works in the process of recognition depends on the rate of success of detection. Human face is very heterogeneous because of the form and skin color. Detection algorithm faces to problems of different lighting scenes, angle of scan camera, head turning and physiognomy. Complexity of the task is proportional to numberless quantity of solutions that arose during research. Methods derive benefit from knowledge and progress from many scientific regions like image processing, probability and statistics, matrix arithmetic, AI, fuzzy logic or spectral analysis.

In present days there are many ways to detect face in an image. On the bases of approach we can divided these methods into four main classes [1].

- **Knowledge-based methods** – The face is determined by defined rules, which describe a “typical face.” Usually these rules describe relations among face features. This method requires very precise localization and description of particular features, which leads to necessity of using complicate and robust algorithms. On that ground these methods in principle do not reach demanded results.
- **Feature invariant approaches** – detection pursuant to generally valid human faces features, which do change to light conditions or head rotation. Such features can be facial features (eyes, nose, etc.), facial texture or skin color.
- **Template matching methods** – searching on the basis of correlation image with preset pattern of either the whole face or its part. A drawback is necessity to create and to keep saved individual pattern in memory. It is mostly necessary to set up these patterns manually, which is very tough and time-consuming.
- **Appearance-based methods** – human face models are obtained by learning from learning set, which contains variety of face patterns. Detection deals with comparing of image segment with the model. There is necessary to create face model again, but the learning is generally semiautomatic. Methods based upon this approach reach very good results.

From see above mentioned reasons the most perspective methods for our work are invariant feature methods.

### 1.1 Feature invariant approaches

A common denominator of these methods is presumption, that there are some features that generally may describe the face independently on its form or background parameters. The methods differ at that way, how to define invariable features and how it is possible to find them in the image. In several methods the invariable features are assumed as a geometric form of each face component part as eyes, eyebrows or mouth. With extraction of these parts by edge detectors and creation of statistical models of their relations, it is then check presence of a face in image.

#### Face features

These techniques are focused directly on searching or facial feature eyes, mouth, nose etc. Various math transformations are used to highlight these features with respect to the rest of face or entire image. Among these techniques belong e.g. edge detection and other filtering methods [2]. More advanced algorithms derive benefit from knowledge of distance of one part from another or from the others. Further these algorithms use probability where we can find particular part and probability where we cannot find it. An example could be the simple fact, that the classifier wouldn't evaluate tested object as the face in that case, in which we can find lips among the eyes and nose. This method fails especially if we find the face in picture with complex environment (high rate of edges and other structural elements).

### Face textures

Human faces have distinct structure, which differs from other objects in image. We usually separate skin, hair and other objects in image. Color information is used to detect face texture [3]. Although this method is fast, it is used rather in more complicated implementations.

### Skin color

Searching for skin-like color regions in color image has become very much effective and used in many applications [4]. The advantage of using this method is rate, above all; farther minimal sensitivity to change luminous conditions, insensitiveness on size and face angular displacement, age, or sex of subject.

## 1.2 Skin detection

Basic principle of all methods coming-out from observation, that the human color skin participate explicit compact subspace in colored space [5]. By the limitations of the space it is possible to model human color skin and then classify the color of each single pixel in image. Skin color classification can be divided into three steps:

- suitable color space selection,
- skin color model selection and the accuracy of this model,
- determination of the criteria for classification.

### Suitable color space selection

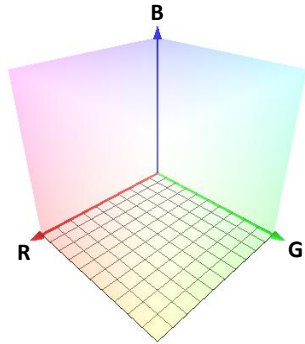
In the field of colorimetry, imaging and image processing, the research has brought many color spaces with various properties. Form and solidity of skin color subspace differs in each model. Suitable color space selection is primary target for many classification methods.

On behalf of fundamental color space used in image processing we can mark color *RGB* space (Fig. 1, *RGB* cube is identical with *RGB* space). The color is defined by three fundamental colors composition (red, blue and green) with various intensities.

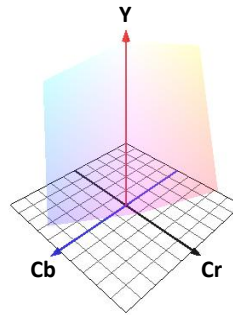
PAL the television standard and JPEG image compress algorithm use *YCbCr* color space. It is transformed *RGB* model, where *Y* represents luminance components; *Cb* and *Cr* are chrominance components (Fig. 2, cube *RGB* is transformed into particular hexahedron). There is a simple relation between both models:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Simplicity and the fact that many formats result from this space assured great popularity for classification systems.



**Fig. 1.** RGB the color space.



**Fig. 2.** YCbCr the color space.

### **Skin color model selection and the accuracy of this model**

The principle of color segmentation in face detection in scene is section individual pixel accordance with its color value into blocks of skin-like pixels and not-skin-like pixels [6]. At the first, it is necessary first to define the color skin area. Explicitly, it is possible to express the color skin area by defined rules, on the basis of distribution skin-like pattern in color space.

The distribution is obtained experimentally by comparison different testing patterns (Fig. 3). The best benefit is the easy implementation, thereby very good classification rate. Quality of results very depends upon chosen model and accuracy of rules, which it is necessity determine experimentally.



**Fig. 3.** Example of skin-like pattern.

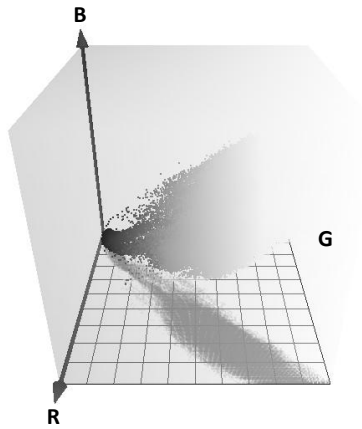


Fig. 4. Skin-like subspace in RGB model.

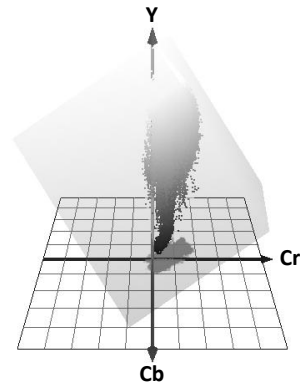


Fig. 5. Skin-like subspace in YCbCr model.

#### *Non-parametric models of skin-like color distribution*

Keynote is to find out skin-like color distribution from training set without explicit expression of color subspace by defined rules. Instead of that it is mostly designed so-called stochastic map – for particular point of discrete color space it is associated probability of that the pixel's color is skin-like color.

#### *Parametric models*

Parametric models try to describe the form of skin-like color distribution with suitable mathematical model. Parameters of the model are then obtained from training set. Skin-like color subspace has in most case elliptic form and normal distribution approximately.

## **2 Automatic face localization algorithms**

On the occasion of relation of the algorithm we went out from presumption that there could occur unlimited amount of faces with different angular displacement and various distances from picture camera. But for localization, we consider those only, in which eyes and mouth are visible and which attain at least minimal size (approximately  $20 \times 20$  pixels). The localization can be divided into two parts. The first part presumably face candidates are looked for and in second there are eliminated inappropriate candidates on the basis of found eyes and lip positions.

## 2.1 Face parts

In this part we search presumptive face region candidates by the use of skin-like color segmentation, wavelet transform and morphologic operations. For skin-like color segmentation we use already created database binary model, which is later converted to stochastic representation.

### Binary pattern database

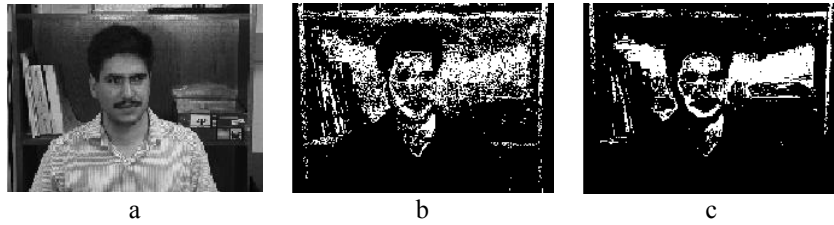
In our case we first assembled sufficient number of different image skin patterns (totally over 400 patterns), which were scanned by three different cameras from various facial parts of 50 people. From this, we created database, which is represented by three-dimension matrix of  $256 \times 256 \times 256$  size, where on index of each matrix element represents corresponding color value of RGB color space and is set either on 1 (given value coincides into skin-like color area) or 0 (given value does not coincide into skin-like color area). Now we are already able to create binary map on the basis of comparing each pixel of tested picture with this database. This map determinates which pixel from tested picture coincides into skin-like color area. The main advantage of using the database, which describes skin-like color region, in comparison with mathematical model [7] is better image processing rate, but although the bigger memory requirements (it is necessary to have saved database in memory, about 16 Mbits) too.

However according to [8] and practical tests, it has been realized, that the RGB color space is not really good for face detection. In practice it is mostly used different color spaces like YCbCr, HSV, RG, etc. For this project it has been chosen YCbCr space as an optimal color space. According to (1) we performed from RGB to YCbCr transform and then operate herewith color space.

### Stochastic pattern database

However we can use the binary database only in conditions that pixels accordant with skin-like color are able to find rarely in the regions out of the face. These pixels also do not flock together. That conditions it is possible satisfy only if the scanned scenes are created artificially, which is practically useless. On this account, we modify the binary matrix onto stochastic matrix with respect to percept occurrence of each color patterns. Further, we then consider skin-like color of face if it is homogenous to the considerable measurement. It means the color value changes gradually only. From here, we proceed on next adjustment of our application. For dispatching of the following processing, we would not proceed pixel-by-pixel, but block-by-block, one of  $3 \times 3$  pixels. For every single pixel of appropriate block we determine probability if the value coincides to skin-like color in accordance with stochastic matrix. On the base of comparison first and second static stochastic moment from all pixels in given block, we assign to this block either the one (high mean value, low scatter – block coincides into skin-like color) or the zero (low mean value, high scatter – block does not coincide into skin-like color). We retrieved binary map again this way, which corresponds to appearance of skin-like color regions. Indeed, this map has one third size only compared to previous map and the percent occurrence of not-skin-like color pixels

coincide into skin-like color region is reduced distinctively. The comparison both binary and stochastic map is in Fig. 6.

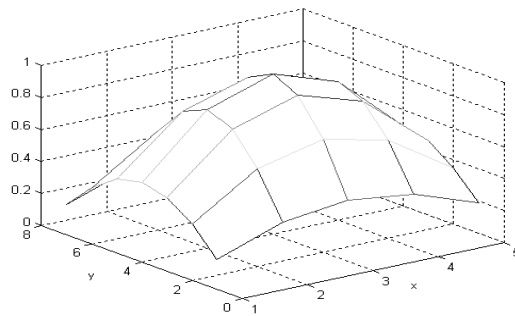


**Fig. 6.** a) original image, b) binary mask in accordance to binary matrix, c) binary mask in accordance to stochastic matrix.

Notice that the amount of not skin-like color pixels is reduced on the other hand amount of skin-like color pixels is reduced too. It is therefore necessary to perform filtration by low-pass filter (for example blurring).

### Wavelet transform

In spite of using stochastic matrix there is significant amount of not skin-like color pixels. To better reduction we treat luminance components of the image to wavelet transform. From presumption there is expressive brightness transition in vertical direction in the area of eyes, lips, etc, we use detailed coefficients in vertical direction and first level of decomposition only. Transitions in horizontal direction are insignificant (especially if the face melts to background). We use Daubeschi wavelet type 2. The high coefficient value means sharp transitions (eyes, lips, etc.). To cover larger face area, we use low-pass filter again. However this time, we go out of presumption that it is most important to distend the coefficients more vertically than horizontally (see face anthropology) and therefore we use special averaged out function of sinusoid (Fig. 7). This field of coefficient we adjust on binary map size and multiply them with this map.



**Fig. 7.** Impulse response of averaged out function of wavelet coefficients.

### Morphologic operations

By the use of wavelet transform we reduced all areas of pixels, which were homogenous in vertical direction. Nevertheless even in the binary map could be still pixels, which do not belong to face area and satisfy all determine criteria so far. Now, we perform series of morphologic operation closing and opening (2) and (3) [9], which result in removing standalone pixels and groupage of other pixels into compact units.

$$X \bullet B = (X \oplus B) \ominus B, \quad (2)$$

$$X \circ B = (X \ominus B) \oplus B, \quad (3)$$

where  $\oplus$  means dilatation morphologic operation (4) and  $\ominus$  erosion morphologic operation (5).

$$X \oplus B = \{p \in \varepsilon^2 : p = x + b, x \in X, b \in B\}, \quad (4)$$

$$X \ominus B = \{p \in \varepsilon^2 : p + b \in X, \forall b \in B\}, \quad (5)$$

Then we spatially describe these units by the use of simple edge detector and suitable tracker algorithm. In our case we described them with rectangular tetragons with  $x_{1i}, y_{1i}, x_{2i}, y_{2i}$ , where  $i$  means order of tetragon and  $x_1, x_2, y_1, y_2$  are initial and end coordinates in horizontal/vertical direction. On the basis of size rate of each side we discard area, which not satisfy form of face [10]. Further we discard such areas, which are too small to provide fair information. Fig. 8 shows all binary masks after averaging, after multiplication with wavelet coefficients and after morphologic operations with suitable face candidate selection.



**Fig. 8.** a) averaged out binary mask, b) multiplied with averaged out wavelet coefficients, c) after morphologic operations with suitable face qualifiers selection.

### 2.2 Face features localization

Indeed, how we see in Fig. 8c, we still have not discarded all areas, which are not skin-like color areas. So we perform last verification test for elected candidates, and it is location the position of mouth and eyes. To solve the problem we accede by the use of so-called color map of each qualifier region. First we focus on lips localization, because this face part is invariable against head turning and pertinent structural face components (beards, glasses, etc.).



### Mouth localization

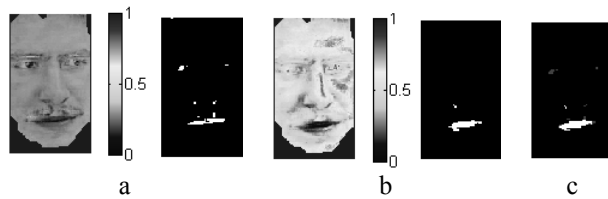
In first part, we perform FLD the color transformation [11] (6), which comes out from presumption that color of lips composition consists of high levels of red and low levels of blue of color space RGB. Unfortunately this color composition varies in various ways in different people and with FLD we are not able to involve all possibilities. From this reason we establish  $rg$  color transformation [12] (7), which would support FLD transformation in some cases.

$$FLD = [-0.289 \quad 0.379 \quad 0.038] \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (6)$$

$$r = \frac{R}{(R+G+B)} \quad g = \frac{G}{(R+G+B)}, \quad (7)$$

We use only  $g$  component, because  $r$  component does not appeared to be good for this, because of redundation and low dispersion. Results are two similar color map of mouth (Fig. 9, we can see very low values in corresponding areas for both cases). Drawback is possibility of occurrence those background parts with very high value of red and low value of blue (for example red scarf about neck, etc.). We use that fact we keep at one's disposal averaged out binary mask of supposed face, thus we process that parts only, which are concerned in this binary mask. Further, there could be that case, in which the red color is in the face in larger quantity (e.g. strongly reddish face), therefore we use already performed wavelet transform in vertical direction (lips appear like sharp transition in vertical direction, especially if opened) and then multiply the result with these maps. Now we transform each map onto binary map by tresholding (thresholds are obtained pursuant to histograms of each color map.). Then we merge these binary maps by (8) and then apply *mean shift* algorithm [13]. The result of this transform is coordinates of area with the best probability of lips appearance.

$$b = \frac{1}{4} \cdot (b_1 + b_2) + \frac{3}{4} \cdot b_1 \cdot b_2. \quad (8)$$

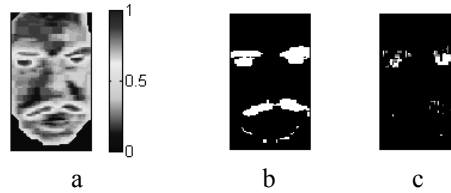


**Fig. 9.** a)  $g$  the color map and its binary image, b) FLD the binary map and its binary image, c) result of (8)

### Eyes localization

By eyes localization we use similar attempts like by the lips localization, however we use only one color map [14] (9). For the reason pupil of the eye and white of the eye colors are very similar in most people. Example of eye color map is in Fig. 10 (areas answer to eyes reach of high level values.). Here again we use averaged out face binary mask because of disturbing background influences (e.g. white shirt collar) and apply multiplication with wavelet coefficients (just in case of closed eye, when there is a influence of eyebrow and eyelashes). And again, we transform color map onto binary mask by the help of tresholding (thresholds are obtained from histogram) and by using mean shifting algorithm we localize eyes position (it is necessary take in team lips position here, because teeth could reach also very high level values in eye color map).

$$b_{eye} = \frac{1}{3} \cdot C_b^2 + (256 - C_r)^2 + \frac{C_b}{C_r}. \quad (9)$$



**Fig. 10.** a) color eye map, b) binary image, c) binary image multiplied with wavelet coefficients.

### 2.3 Face regions verification

On the base of mutual positions and distances among lips, both eyes with a respect to vertical and horizontal size supposed face we determine whether it is face or not [10]. In the case of agreeable the face coordinates are saved including the size and eyes and lips positions for next appropriate process. The face is included into result image. In opposite case the qualifier areas are reprobate. All qualifier areas are preceded with this detection by sequel. The final result is in Fig. 11 (green dots mean approximate eyes and lips positions).



**Fig. 11.** a) original image, b) detected face with eyes and lips highlight.

### 3 Conclusion

For benchmarking developed algorithm, we have chosen *Georgia Tech face database* [15], which has 750 shots of 50 subjects (men and women in different ages). The advantage compared to others is that it includes complex backgrounds in combination with skin-like color background. This presents challenge for many face localization algorithms. Another advantage is subject diversity, which are various head turning, emotional expression, lightning and structural face accessories (beard, glasses, etc.). We achieved 93,4% rate of success, which is very good result with regard to complexity of the database. Main cause of unsuccessful localization was especially insufficient suppression of color influence of background in the area of face with wavelet coefficients. It would be good to find better way to separation in next work in the future. To other problem was present of dense beard, in which lips was not included to facial area and face candidate had to be then denied. This would be prevented by vertical expansion of area in lips localization even except candidates face area.

**Acknowledgements.** The paper was prepared within the framework of No. 102/07/1303 project of the Grant Agency of the Czech Republic and No. 1ET301710509 project of the Czech Academy of Sciences. The paper has been supported by the Ministry of Education, Youth and Sport of the Czech Republic project FRVS 972/2007/G1.

### References

1. Yang, M.-H.: Detecting Faces in Images: A Survey. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 24, 1. (2002)
2. Yow, K. C., Cipolla, R.: Feature-Based Human Face Detection, *Image and Vision Computing*, Vol. 15. (1997)
3. Augusteijn, M. F., Skujca, T.L.: Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology, *Proc. IEEE Conf. Neural Networks*. (1993)
4. Jones, M. J., Rehg, J.M.: Statistical Color Models with Application to Skin Detection, *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Vol. 1. (1999)
5. Graf, H. P., Chen, T., Petajan, E., Cosatto, E.: Locating Faces and Facial Parts, *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*. (1995)
6. Singh, S. K., Chauhan, D. S., Vatsa, M., Singh, R.: A Robust Skin Color Based Face Detection Algorithm, *Tamkang Journal of Science and Engineering*, Vol. 6. (2003)
7. Garcia, C., Zikos, G., Tziritas, G.: Face Detection in Color Images using Wavelet Packet Analysis, *IEEE Int. Conf. on Multimedia Computing and Systems*. (1999)
8. Kawato, S., Ohya, J.: Automatic Skin-color Distribution Extraction for Face Detection and Tracking, *ICSP2000: The 5th Int. Conf. on Signal Processing*, Beijing, China. (2000)
9. Boomgaard, Balen: Image Transforms Using Bitmapped Binary Images, *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*. (1992).
10. Young, J. W.: *Head and Face Anthropometry of Adult U.S. Civilians*, U. S. Government Printing Office. (1993)

11. Chaloupka, J.: Rozpoznávání akustického signálu řeči s podporou vizuální informace, disertační práce Technická univerzita v Liberci. (2005)
12. Stern, H., Efros, B.: Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments, Proceedings of the fifth IEEE International Conference on Automatic Face and Gesture Recognition. (2002)
13. Toth, D., Stuke, I., Wagner, A., Aach, T.: Detection of Moving Shadows using Mean Shift Clustering and a Significance Test, Proceedings of the 17th International Conference on Pattern Recognition. (2004)
14. Campadelli, P., Cusmai, F., Lanzarotti, R.: A Color-Based Method for Face Detection, Proceedings of the International Symposium on Telecommunications (IST2003), 186-190. (2003)
15. Nefian, A., V.: Georgia Tech face database, [http://www.anefian.com/face\\_reco.htm](http://www.anefian.com/face_reco.htm)