

Evaluation of RGB and HSV models in Human Faces Detection

Marián Sedláček

sedlacek.marian@pobox.sk

Faculty of Informatics and Information Technologies

Slovak University of Technology

Bratislava / Slovakia

Abstract

This paper presents detection of human faces in a color image. The detection is based on a skin-color model represented by a Gaussian model. We compare 12 different skin-color models that vary in aspects of color representation of a pixel (RGB, HSV, HSL), complexity of Gaussian model and character of an input image set. We present all steps of the image processing and some assumptions to optimise the results. Finally, conclusions are presented and future work is outlined.

Keywords: skin-color model, face detection, Gaussian model

1 Introduction

In the past decade, face detection has become an often researched problem. It is the primary step of other tasks such as face tracking in color video sequences or recognition of facial features. This research area has many applications in face identification systems, model-based coding, gaze detection, teleconferencing, augmented reality, etc. It also helps to solve the idea of simple human-computer interaction and communication [1].

The face detection systems have to detect every human face in an input image, no matter the lighting conditions or race of people in the image. These systems are usually based on an experimentally estimated skin-color model. Skin-color model uses an idea that color distribution of skin-color of different people is clustered in a small area of chromatic space.

The main goal of our work is create a face detection system by using a statistical skin-color model represented by Gaussian model. We will analyze 12 different skin-color models that vary in aspects of color representation of a pixel (RGB, HSV, HSL), complexity of Gaussian model and character of an input image set.

In Section 2, we describe the present state of methods used in the area. In Section 3, we discuss difficulties and describe techniques which we subsequently developed to overcome them. Finally, in Section 4, we present the results of our experiments.

2 Related works

Many research results on automatic face detection have been published. The well-known method is using an empiric skin-color model that is modelled by a single Gaussian model. It is based on a fact that distribution of skin-color of different people can be represented by a Gaussian model. Using the single Gaussian model is fast. But it does not adequately represent the variance of this skin distribution occurred in the situation where illumination condition varies. To overcome this drawback, we can use a finite Gaussian mixture model whose parameters might be estimated through the Expectation-Maximization (EM) algorithm. [2,3]

Another technique is using adaptive histogram backprojection. An initial estimate of the skin model is the 2-D histogram $S(r,g)$ obtained from cut-out skin regions in the face. The frame to be segmented is transformed into rg-space and each pixel p_i with chromaticity (r_i, g_i) is assigned value of the histogram at (r_i, g_i) , $S(r_i, g_i)$. A variation is to use a ratio histogram $R(r,g)$, which is $S(r,g)$ by the whole image histogram $I(r,g)$ to penalize colors which are not part of the model or are present also in the background, thus increasing the contrast between skin and background pixels. With ratio histogram and histogram backprojection, no fitting (e.g. Gaussian) is necessary because the histogram itself is used as the model, and probabilities are assigned by simple table lookup, thus leading to a faster labeling. But it becomes effective only when training data is sufficiently large to be dense. Moreover, it requires additional memory to keep the histograms. [4]

Another technique is using an elliptical boundary model. It is based on an observation, that the skin area in each chrominance space fits well an ellipse. This model is trained from a set of training data in two steps, preprocessing and parameter estimation. In preprocessing step outliers are removed so that the trained model reflects the main density of the underlying data set. In parameter estimation step are estimated model parameters from the preprocessed data set. [4]

3 Face detection

3.1 Skin-color models

The most important part of this project was to find an appropriate skin-color model. The skin-color model should be adaptable for any skin color at any lighting conditions. The common RGB representation of color images is not suitable for characterizing skin-color. In the RGB space, the triple component (r, g, b) represents not only color but also luminance. Luminance may vary across a person's face due to the ambient lighting and is not a reliable measure in separating skin from non-skin region. Luminance can be removed from the color representation in the chromatic color space. Chromatic colors, also known as "pure" colors in the absence of luminance, are defined by a normalization process shown below:

$$r = R/(R+G+B) \quad (1)$$

$$g = G/(R+G+B) \quad (2)$$

The normalized blue color is redundant because $r + b + g = 1$.

Skin colors of different people appear to vary over a wide range, they differ much less in color than in brightness [3]. So, the colors of human skin fit in a small area of chromatic color space. In the following section, we describe the process of estimation of our skin-color model.

We collected two sets of 15 color images each with human faces from the World Wide Web. First set are images of people with white colored skin (Caucasian and a part of Asian race), second set of people with brown and black colored skin (African and a part of Asian race). Then we manually selected little rectangle samples of skin from every image of each set. These samples were filtered using a low-pass filter to reduce the effect of noise. Then we counted normalized values of red and green color for each pixel of filtered samples (formulas 1, 2).

As shown in Figure 1 distribution of skin-color of different people is clustered in a small area of chromatic space and can be represented by a Gaussian model. Gaussian model $N(m, C)$ is a kind of normal statistical model that is estimated with parameters - mean vector and covariance matrix:

$$\text{Mean: } m = E \{ x \} \text{ where } x = (r \ g)^T \quad (3)$$

$$\text{Covariance: } C = E \{ (x - m)(x - m)^T \} \quad (4)$$

Assuming that the skin color density is modelled by a Gaussian model, the skin likelihood of an input chrominance vector x is given by formula :

$$p(x) = \exp [-0.5 (x - m)^T C^{-1} (x - m)] \quad (5)$$

where $x = (r \ g)$, m is mean vector, C is covariance matrix.

So, finding an appropriate skin-color model depends on estimating right parameters of Gaussian model. The main aim of our project was to compare processing of images with different skin-color models.

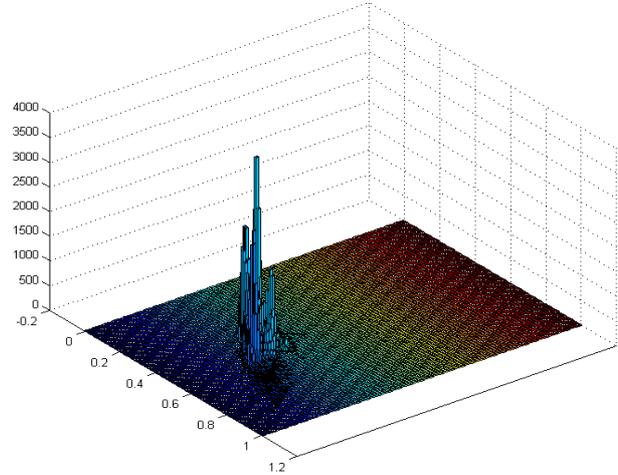


Figure 1: Color distribution for skin-color of different people

Skin-color models based on Gaussian model can vary in these aspects:

- character of an input image set
- color representation of a pixel
- complexity of Gaussian model

As was said in the beginning of this section, we had two sets of color images. So we could create three sets for analyzing: one set of people with white skins (*Set W*), second set of people with black (*Set B*) and brown skins and third set as an union of first and second set (*Set WB*). We used following color representation of a pixel in our experiments: (normalized) RGB, HSV and HSL. Estimating parameters of skin-color model using HSV and HSL model is similar to the (normalized) RGB model. The main difference was that every component of HSV representation (h, s, v) was relevant, so the mean vector and covariance matrix of Gaussian model were 3-D.

The complexity of Gaussian model divides models in two basic types – single Gaussian model and mixture Gaussian model. The parameters of mixture Gaussian model can be estimated by means of the Expectation-Maximization (EM) algorithm [4].

Because of the complexity of EM algorithm we estimated the weights of each Gaussian model experimentally and we will calculate skin likelihood of an input chrominance vector x by formula :

$$p(x) = 0.3N_W(m, C) + 0.4N_B(m, C) + 0.3N_{WB}(m, C) \quad (6)$$

where $x = (r\ g)$, $N_W(m,C)$ is Gaussian model of Set W , $N_B(m,C)$ is Gaussian model of Set B , $N_{WB}(m,C)$ is Gaussian model of Set WB .

Accordingly we had 12 different Gaussian distributions to compare: 1-3: Single Gaussian Model based on RGB color representation and Sets W , B , WB , 4-6: Single Gaussian Model based on HSV color representation and Sets W , B , WB . 7-9: Single Gaussian Model based on HSL color representation and Sets W , B , WB , 10-12: Mixture Gaussian Model based on RGB, HSV, HSL color representation.

3.2 Skin-Likelihood Image

The first step in the processing of an input picture is creating a skin-likelihood image. Skin-likelihood image is an image in which each pixel corresponds to the probability of occurrence of skin-color (in the same pixel in the original input image). The probability of each pixel is calculated by formula (5). The values of probability can be easily transformed into greyscale values. So skin regions are brighter than the other parts of image.



Figure 2: Original image, skin-likelihood image

Note: To reduce the effect of noise in an input image is useful to use a low-pass filter. See section 4.3.

3.3 Skin-Segmented Image

The second step is creating a skin-segmented image by using a threshold value of probability. If the probability of a pixel in skin-likelihood image is more or equal to estimated threshold value, we suppose that this pixel represents skin color. If not, we suppose that this pixel does not represent skin color. The skin color pixels are white and the other ones are black in skin-segmented image.

Estimating a threshold value is very important for next steps of image processing. We can use fixed threshold value for every image or adaptive thresholding. The adaptive thresholding is based on the observation that decreasing the threshold value may intuitively increase the segmented region.

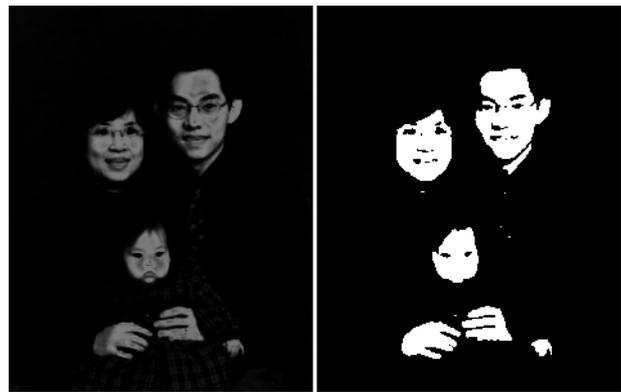


Figure 3: Skin-likelihood image, skin-segmented image

However, the increase in segmented region will gradually decrease, but will increase sharply when the threshold value is too small that other non-skin regions get included. The threshold value at which the minimum increase in region size is observed while decreasing the threshold value will be the optimal threshold. [3]

We found out that using a fixed threshold value is more efficient in our experiments. However, we implemented both ways of thresholding process in our program and so user can easily choose which one he wants to use.

3.4 Selection of face regions

Using the result from the previous section, we proceed to determine which regions can possibly determine a human face. We will consider following assumptions that were obtained in our experiments (the *Assumption A* is published in several articles):

- A. A human face is defined as a closed region in the image, which has 1 or more holes (eye, mouth, etc.) inside it.
- B. The ratio of width and height of a human face is not bigger than 3.0.
- C. A segmented human face region is not 5 times smaller than the maximal square of all segmented regions that verify *Assumptions A* and *B*.

The idea of our algorithm for *Assumption A* is as follows. We are looking for a closed white (skin) region that has one or more black (not-skin) regions inside. In other words, we are looking for a black region that is bounded with a white region. Accordingly, for every pixel of the black region, following rule must be true. If we move from the pixel to the left, to the right, up and down, we should find 4 pixels that are part of the same white region. If this rule for every pixel of a black region true, it means that this region is bounded by a white region. In other words, the white region has a black hole inside.

To make this algorithm more simply and its execution faster, we assume that there is no such white region, that has a black region inside and there are also another one

or more white regions inside that black region. With this assumption we can stop searching for a white pixel in that 4 directions as soon as we found first white pixel. We can make this reduction because, we are looking for black holes which results from facial features such a mouth, eyes and there is very low probability that for example inside a human mouth is something that has skin-color.

Before we start to find the white regions that has one or more black holes inside, we need to label all white and black regions with an unique label. We used unique colors as labels and 8-connected seed fill algorithm for labelling of all white regions as well as 4-connected seed fill algorithm for all black regions.



Figure 4: Skin-segmented image, skin-segmented image with white regions labelled



Figure 5: Skin-segmented image with white and black regions labelled, skin-segmented image with selected skin regions applying *Assumption A*

Assumption B tells that the ratio of width and height of a human face is not bigger than 3.0. Usually the ratio of width and height of a normal human face is smaller, but human faces can have different orientation in a image and sometimes we detect face region as face together with neck. We can see a positive example of an application of *Assumption B* at Figure 6 where a part of the image has skin-color although it is not a part of human body. *Assumption C* tells that a segmented human face region is not 5 times smaller than the maximal square of all segmented regions that verify *Assumptions A* and *B*. It might happen that applying only *Assumption A* and *B* is not enough to segment face regions.



Figure 6: Original image, skin-segmented image with selected skin regions applying *Assumption A*, result image applying *Assumptions A* and *B*

For example, hands have skin-color, segmented region of hands can have a hole inside (between fingers) and the ratio of its width and height is not bigger than 3.0. So it verifies *Assumptions A* and *B*. We can see a positive example of an application of *Assumption C* at Figure 7 where a little region of hands is not included in the result image.

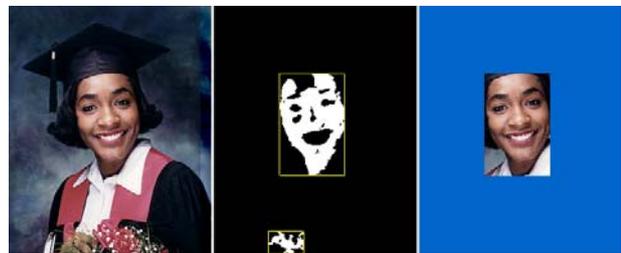


Figure 7: Original image, skin-segmented image with selected skin regions applying *Assumption A*, result image applying *Assumptions A, B* and *C*

The selection of face regions has usually better results with applying these 3 assumptions as without them. Especially, if the input image is in a good quality and has a portrait character. But sometimes it might happen that on of the assumptions makes a result image wrong. Therefore user can enable or disable applying of any assumptions.

Note: There are some recommendations about applying these assumptions in section 4.3.

4 Results

4.1 Evaluation of the proposed skin-color models

In this section, we compare all 12 skin-color models to select 4 of them that are the most relevant for further comparasion. We collected a set of 8 images of people of different races and classified the quality of each result in scale 0-10 points (mark), where 10 points is the best possible result. We took into consideration the quality of skin-region segmentation and appearance of skin-colors in the background of processed image in this

classification. It has been done by more people and the final value was the average of their subjective valuation.

Skin-color model	Average mark [points]	Skin-color model	Average mark [points]
SG/rgb/WB	8.0	SG/rgb/B	8.4
SG/hsv/WB	7.75	SG/hsv/B	8.0
SG/hsl/WB	7.5	SG/hsl/B	8.0
SG/rgb/W	8.66	MG/rgb	7.875
SG/hsv/W	9.66	MG/hsv	8.5
SG/hsl/W	9.66	MG/hsl	8.125

Figure 8: Table of primary results (*SG* is single Gaussian model, *MG* is mixture Gaussian model)

Skin-color models (*SG/rgb/W*, *SG/hsv/W*, *SG/hsl/W*) based on *Set W* where tested only with images of people having white skin. Skin-color models (*SG/rgb/B*, *SG/hsv/B*, *SG/hsl/B*) based on *Set B* where tested analogue. The aim of this project was to estimate a skin-color model that would be adaptable for any skin color and so these 6 models were not relevant for further comparison.

As shown at Figure 8 the most 5 best skin-color models are: 1. *MG/hsv*, 2. *MG/hsl*, 3. *SG/rgb/WB*, 4. *MG/rgb*, 5. *SG/hsv/WB*. The testing set was too small to see objective results, so we will continue the comparison. According that the results of HSV and HSL skin-color models are almost the same, we will ignore the HSL models.

4.2 Final results

According the results from previous section, we will compare following skin-color models: *MG/hsv*, *MG/rgb*, *SG/hsv/WB*, *SG/rgb/WB*. We collected 3 new sets of images: *Set CAU* of 11 images (Caucasians), *Set ASI* of 9 images (Asians) and *Set AFR* of 10 images (Africans).

Skin-color model	Average mark [points]	Skin-color model	Average mark [points]
SG/rgb/WB	7.18	MG/rgb	7.72
SG/hsv/WB	7.36	MG/hsv	8.00

Figure 9: Table of final results by using *Set CAU*

Skin-color model	Average mark [points]	Skin-color model	Average mark [points]
SG/rgb/WB	6.77	MG/rgb	7.77
SG/hsv/WB	7.88	MG/hsv	8.77

Figure 10: Table of final results by using *Set ASI*

Skin-color model	Average mark [points]	Skin-color model	Average mark [points]
SG/rgb/WB	7.10	MG/rgb	7.00
SG/hsv/WB	7.00	MG/hsv	7.00

Figure 11: Table of final results by using *Set AFR*

Skin-color model	Average mark [points]	Skin-color model	Average mark [points]
SG/rgb/WB	7.01	MG/rgb	7.50
SG/hsv/WB	7.41	MG/hsv	7.92

Figure 12: Table of final results by using all *Sets CAU, ASI, AFR*

As shown at Figure 12 the *MG/hsv* skin-color model appears to be the best for any skin color. Using mixture Gaussian models is generally more effective than using the single ones, as well as using HSV (or HSL) color representation than RGB.

4.3 Interesting aspects of processing

Because of it is allowed to enable or disable mentioned 3 assumptions or pre-filtration in our software, there can be different results for the same input image and skin-color model. In this section, we will analyze some interesting aspects of image processing.

The pre-filtration is activated as a default option. As shown at Figure 13 segmented skin regions are more integral if we use low-pass filter to reduce the effect of noise.



Figure 13: Image processing applying pre-filtration and Assumptions A, B, C



Figure 14: Image processing without pre-filtration and assumptions A, B, C

Using pre-filtration and so making skin segmented regions more integral might be also a disadvantage. For example, if the ratio of squares of image sizes and maximal segmented skin region is more than 30:1, there need not to be any black hole from a facial feature segmented. Accordingly applying *Assumption A*, we might ignore some truly skin regions (see Figure 15).

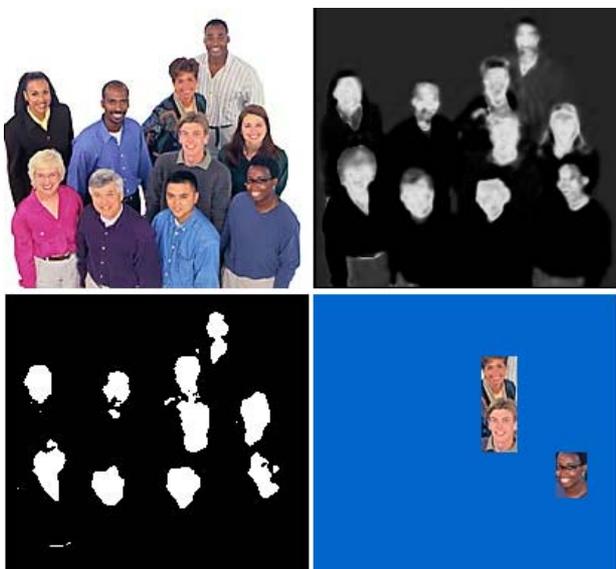


Figure 15: Image processing applying pre-filtration and *Assumptions A, B, C*

So, if the ratio of squares of image sizes and maximal segmented skin region is more than 30:1, we have two options to optimize the result. First is to disable pre-filtration (see Figure 16), second is to disable *Assumption A*.

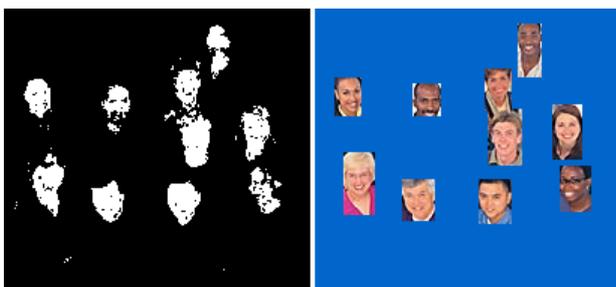


Figure 16: Image processing without pre-filtration and *Assumptions A, B, C*

The *Assumption A* is also good to disable, if a face on an input image has side orientation. If we process an image where people have no dress (the neck and upper part of the body), we should disable *Assumptions B* and *C*. If we want to characterize the main difference between using RGB and HSV color representation, we can say that the skin-color models using RGB representation has more variance. In other words, it detects more hues of skin-colors than the HSV one. This aspect can be effectively used in images of faces that have some parts more affected with ambient light.



Figure 17: Image processing using *MG/rgb* skin-color model



Figure 18: Image processing using *MG/hsv* skin-color model

There are cheeks and jaw more lightened than other parts of the face at the Figure 17. Skin-color model *MG/rgb* segments even the more lightened parts. But as we can see at Figure 18, the *MG/hsv* model does not.

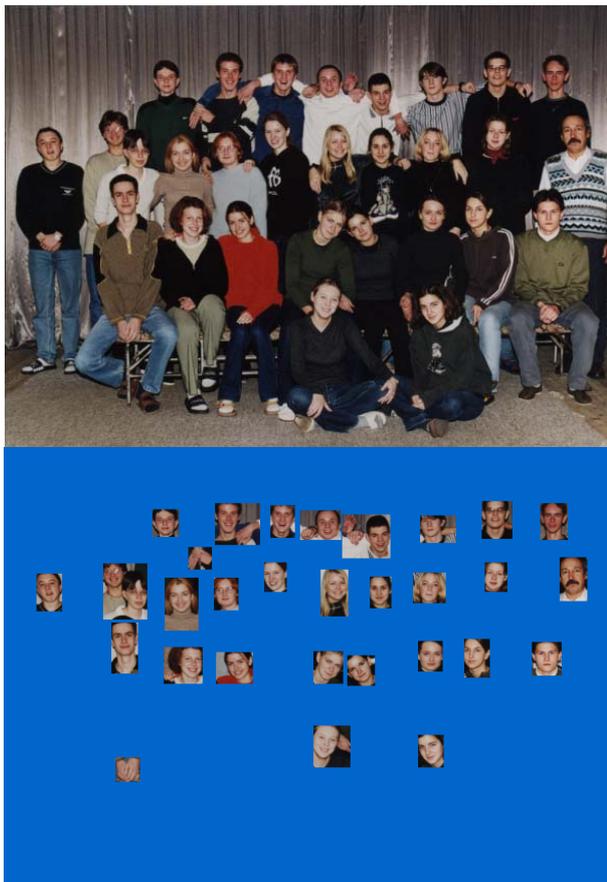


Figure 19: Image processing using *MG/hsv* skin-color model, *Assumptions A, B, C* and without pre-filtration

Figure 20 illustrates the duration of this computing. It depends on the sizes of the image, number of segmented regions and square of black holes from facial features.

Image at	Size [pixels]	Duration [seconds]
Figure 2	338x427	28
Figure 6	163x253	14
Figure 15	204x190	10
Figure 19	600x432	117

Figure 20: Duration of image processing at computer AMD Duron 990MHz, 256MB RAM

5 Conclusion

In this paper, we presented a method for the detection of human face in a color image. It uses Gaussian models to represent skin-color models. It is evident, both from the histograms of samples and the results, that a Gaussian mixture is more appropriate than a single Gaussian function in estimating the distribution of skin color. We compared evaluation of different color representation and found out that HSV model is better than RGB one. We suggested 3 assumptions to optimize the final result of segmentation and gave recommendations about using of them. To improve the quality of estimated skin-color models, we should use significantly larger sets of analyzed samples and EM algorithm to estimate more appropriate values of mixture Gaussian models.

The evaluation method of processed images is subjective. We plan to use a metric based on automatic comparison in our next work.

Our experiments resulted in the observation how to get more accurately segmented face regions. It is based on the fact that the results of images of people having white skin are better, if we use a skin-color model based only on the *Set W* (analogue by images of people having black skin). So at first, we will use a mixture Gaussian model to get rectangle regions of human face (as in result images at Figures 6, 7, 13, 14, 15,...,19). Second, we will analyze only these regions with a skin-color model based on *Set W* and a skin-color model based on *Set B* separately. Then we will use the one of them which gives bigger average probability of segmented (white) skin regions as the final result.

6 Acknowledgements

This work was a semester project in the subject Computer Graphics 2 at FIIT STU. I would like to thank my professor Martin Šperka for an inspiration, suggestions in the research and help with this paper. Also thanks to colleagues Michal Slamka and Erik Štetina for math-lab scripts to display histograms (Figure 1) and compute mean vector and covariance matrix of a dataset.

References

- [1] Gejuš P., Šperka M., *Face tracking in color video sequences*, Proceedings of SCCG 2003, pp. 268-273, Budmerice, Slovakia, April 2003
- [2] Yang M.-H., Ahuja N., *Gaussian Mixture Model for Human Skin Color and its Applications in Image and Video Databases*, In the 1999 SPIE/EI&T Storage and Retrieval for Image and Video Databases, pp. 458-466, San Jose, January 1999, <http://www.dcs.ex.ac.uk/people/wangjun/yang99gaussian.pdf>
- [3] Chang H., Robes U., *Face detection*, May 2000, <http://www-cs-students.stanford.edu/~robles/ee368/main.html>
- [4] Lee J.Y., Yoo S.I., *An Elliptical Boundary Model for Skin Color Detection*, The 2002 International Conference on Imaging Science, Systems, and Technology, Las Vegas, USA, June 2002, <http://ailab.snu.ac.kr/publication/download/CISST02-169CT.pdf>
- [5] Jones M. J., Rehg J. M., *Skin Color Modeling and Detection*, Hewlett-Packard Company, Hewlett-Packard Company, June 2002, <http://crl-download.crl.hpl.hp.com/vision/humansensing/skin/default.htm>
- [6] Caetano, T. S., Barone, D.A.C., *A Probabilistic Model for the Human Skin Color*, Proceedings of ICIAP2001 - IEEE International Conference on Image Analysis and Processing, pp. 279-283, Palermo, Italy, September 2001, <http://www.cs.ualberta.ca/~7Etcaetano/iciap2001.pdf>
- [7] Kawato S. and Ohya J., *Automatic Skin-color Distribution Extraction for Face Detection and Tracking*, ICSP2000: The 5th Int. Conf. on Signal Processing, vol.II, pp.1415-1418, August 2000, Beijing, China, <http://www.mis.atr.co.jp/~skawato/pdfs/ICSP2000.pdf>