

Dynamic Skin Detection in Color Images for Sign Language Recognition

Michał Kawulok

Institute of Computer Science, Silesian University of Technology, Akademicka 16,
44-100 Gliwice, Poland
michal.kawulok@polsl.pl

Abstract. Skin detection is the first step of processing in many approaches to face and gesture recognition. This paper presents research aimed at detecting skin in digital images for Polish Sign Language recognition. There are many methods for detecting human skin, including parametric skin models defined in various color spaces and statistical approaches which require appropriate training. The presented method is based on statistical model updated dynamically for every image in which human faces can be detected. The detection is performed in luminance channel based on geometric properties of human faces. The experiments proved that effectiveness of this approach is higher than application of general skin detection models.

1 Introduction

Sign languages were developed in communities of deaf people and they are the main means of communication among them. Information is conveyed by gestures and lips movement instead of sound as in the case of natural spoken languages. Noteworthy, the sign languages developed independently in many communities and there is no universal one. Similarly as in the case of spoken languages, the sign languages vary from one geographic region to another, however the coverage of their usage is often not identical to that of spoken languages. For example, the British Sign Language is very different from the American Sign Language (ASL), which is used not only in the area of the US, but also in Mexico and Canada. The research presented in this paper is the first step of a project focused on Polish Sign Language which is used in Poland.

The main problem that appears nowadays is that the majority of people cannot use and do not understand sign languages which creates a barrier separating deaf communities from the rest of society. This barrier can be reduced or even eliminated by automatic systems which translate between spoken and sign languages. A crucial part of these systems is detection and recognition of gestures in digital images. The research presented in this paper is focused on detecting skin areas in order to determine position of faces and hands.

Skin detection is a widely investigated area and many methods for solving this problem have been developed [1, 2, 4, 5, 6, 8, 10, 13, 14]. They can be divided into parametric and statistical methods which are described in section 2. Some

of these methods are quite effective, but they are highly sensitive to changes of lighting conditions and they also produce high false positive errors in case of objects which color is similar to human skin.

Face detection has many applications ranging from entertainment to surveillance tracking and it is also a preliminary stage of face recognition [15]. There are approaches to face detection, which utilize color information to identify skin regions at first and then to build facial feature maps [6]. For frontal face images the detection can be successfully performed in luminance channel based on characteristic properties of face geometry.

The goal of this research was to detect faces and hands which are crucial in case of sign language recognition. At first, faces are detected based on luminance channel. It is feasible, because they are usually oriented frontally to the camera in the analyzed images. After that skin areas are detected using statistical skin color model in RGB color space, which is updated dynamically based on detected face region. The experimental results presented in section 5 confirmed that the developed approach is more effective than other skin detectors.

2 Color Based Skin Detection

It was observed [6,13] that skin-tone color has common properties and that it can be defined in various color spaces after applying color normalization. In general, skin detection methods are based on parametric and statistical skin modeling. The main problem concerned with skin color models is their sensitiveness to changes in lighting conditions and high false positive rates in case of objects which color is similar to human skin.

2.1 Parametric Skin Models

Parametric skin models [4,6,10,11,14] are based on fixed decision rules defined empirically in various color spaces after analysis of skin-tone distribution. Those rules are applied to determine if a pixel color value belongs to skin or not. In this research four parametric models in different color spaces were investigated.

Kovac et al. [10] proposed a model defined in RGB color space. Skin-tone color has been also modeled in HSV space by Nikolaidis [14]. These both models specify conditions for pixel color values and if they are met, a pixel is classified as skin. The models are applicable for various illumination conditions and do not require any color normalization.

An approach proposed by Hsu et al. [6] takes advantage of common skin color properties in nonlinearly transformed YC_bC_r color space. At first, the top 5% of the luminance values in the image define a *reference white* color and the pixel values are scaled linearly in RGB space, so that all reference-white pixels have values equal to 255. After the normalization, skin is detected based on an elliptical model defined in nonlinearly transformed YC_bC_r space.

There are also techniques which operate in two color spaces to increase detection accuracy. A composed skin color detector proposed in [11] is defined in RGB and YC_bC_r color spaces.

2.2 Statistical Skin Modeling

Parametric methods are aimed at finding models of skin-tone color in a given color space. It is also possible to analyze distribution of skin pixel values for a given training set of images, in which skin and non-skin areas are marked. This allows to create a statistical model, which determines probability that a given pixel value belongs to the skin class. Several methods were developed [2, 5, 8] which implement such approach to skin detection.

At first, based on the training set, histograms for skin (C_S) and non-skin (C_{NS}) classes are built. For every color value (v) probability of observing it in each class is calculated:

$$P(v|C_x) = C_x(v)/N_x, \quad (1)$$

where $C_x(v)$ is number of v -color pixels in x -th class and N_x is total number of pixels in that class. Maximal number of histogram bins depends on pixel bit-depth and usually it is $256 \times 256 \times 256$. However, the experiments showed that detection is more effective if number of bins is reduced to about 64 per channel.

To detect skin with low false positive error it is not only important that a pixel value appears frequently as skin, but also its density in the non-skin class should be low. Hence, probability that a given pixel value belongs to the skin class ($P(C_S|v)$) is calculated based on Bayes rule:

$$P(C_S|v) = \frac{P(v|C_S)P(C_S)}{P(v|C_S)P(C_S) + P(v|C_{NS})P(C_{NS})}, \quad (2)$$

where $P(C_S)$ and $P(C_{NS})$ are estimated based on number of pixels in both classes. If the training set is big enough, probabilities for all possible color values can be estimated to create skin color probability map. However, if a value is not observed in the training set, it is assumed that probabilities $P(v|C_S)$ and $P(v|C_{NS})$ equal 0.5. After obtaining the probability map, every pixel value can be transformed into probability that the pixel belongs to the skin class. Skin regions are extracted based on acceptance threshold.

3 Face Detection

Face detection [6, 15] is aimed at finding location of human faces in digital images. If faces are oriented frontally, which can be assumed in case of sign language recognition, their position is unambiguously defined by central points of eyes. There are many techniques of face detection which utilize common geometrical features of human faces or properties of skin color.

In the presented research face detection was performed in luminance channel based on geometric properties of human faces. The first stage is a preliminary selection of *face candidates*, which are those areas in images where faces may be located. After that, the candidates are verified and those of them which are accepted are regarded as faces (Fig. 1). This approach is effective and faces are detected with high precision [9].

Usually human heads have shape of vertically oriented ellipses, so at first such ellipses are detected in images using Hough transform [7]. After that, small horizontally oriented ellipses are detected inside upper half of every ellipse to find candidates for eye sockets. The eye-socket candidates are verified and if there are two positive responses inside an ellipse, a new face candidate is created.

The candidates are verified with the Support Vector Machines (SVM) [3] which are robust learning machines solving two-class classification problems. In order to use the SVM as a verifier, it is trained with two classes of samples: normalized face images of constant size in which eyes are in invariant positions and normalized non-face candidates. Similarly, the SVM is trained and used for eye sockets verification.

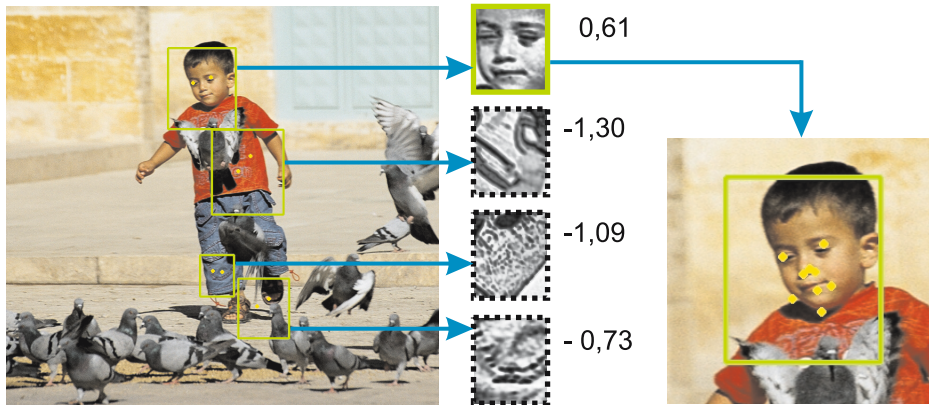


Fig. 1. Example of face detection. Face candidates are detected (*left*), normalized and verified with the SVM (*middle*). If its response (*shown for every normalized image*) is positive, a candidate is regarded as a face.

4 Face-Based Dynamic Skin Model

Skin detection algorithms are vulnerable to lighting condition changes and they generate high false positive errors in case of objects which color is similar to human skin. In vast majority of cases lighting conditions are constant within an image and skin color do not vary much for a single person. Therefore, if skin color could be defined separately for every person, the errors should be significantly decreased. This can be achieved if it is known that a given region of the image contains skin pixels. Based on that region a statistical model can be built or conformed dynamically and applied to the whole image.

If a face is detected in an image, it is possible to point a region which with high probability contains only skin pixels. In the proposed approach such face region is defined based on detected positions of eyes as presented in Fig. 2.

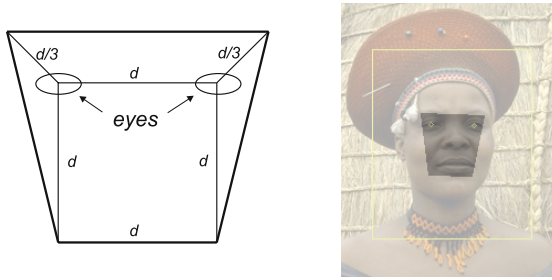


Fig. 2. Face region generation based on positions of eyes

Pixels inside the face region are regarded as skin pixels and probability of observing a given value within the region ($P(v|C_S)$) is calculated (1). The distribution of non-skin pixel values is unknown and it was estimated that number of non-skin pixels for a given value is:

$$C_{NS}(v) = \max_i(C_S(i)) - C_S(v) , \quad (3)$$

which makes it possible to build the dynamic skin-color probability map (2) based on detected face region. As mentioned in 2.2, optimal number of histogram bins for a general model is 64 per channel. In the presented case only a part of color domain is covered by pixels from the face region and it occurred to be more effective to create the face-based probability map for 16 bins per channel.

A model created in this way takes into account only distribution of skin pixel values, which allows for detecting skin areas effectively, but in some cases false positive error is high. The results occurred to be much better if the face-based model was joined with a general statistical model trained with a large training set as described in section 2.2. Both models are defined by probability maps (P_{face} and P_{gen}) which are merged to obtain the final probability map:

$$P(C_S|v) = \omega P_{face}(C_S|v) + (1 - \omega) P_{gen}(C_S|v) , \quad (4)$$

where ω is a weight of the dynamic model. Based on experiments its value was set to 0.8. Probability maps must have the same pixel value domain, so the face-based map is converted from 16 to 64 bins per channel. If no face is detected in an image, the general model is applied to detect skin regions. Example of skin detection with the described approach is presented in Fig. 3. The background color is similar to human skin which results in very high false positive rate (68.9%) for a general model. False positives were reduced to only 1.1% when the model was merged with the dynamic face-based skin model.

5 Experiments

The experiments were conducted for ECU face detection database [12] which contains 4000 images with marked areas of human skin. The images were divided into two groups: a testing set, for which the detection performance was

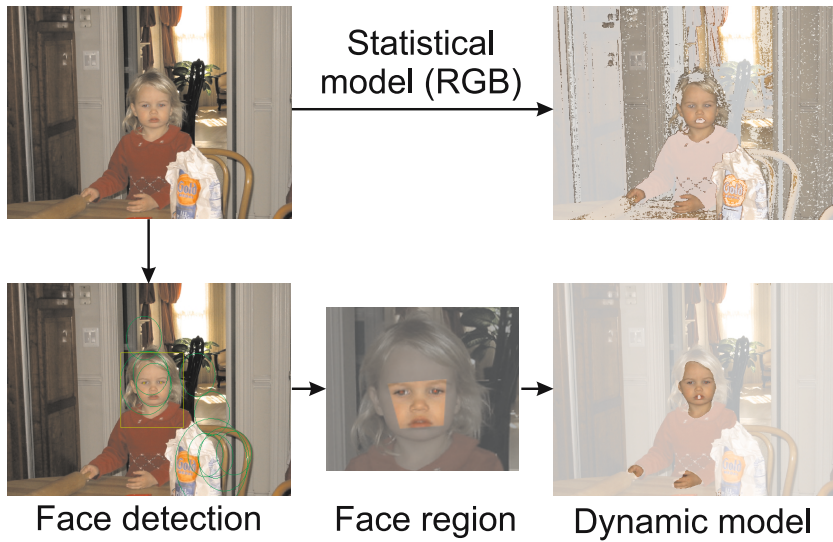


Fig. 3. Skin detection with a statistical model and with a face-based dynamic model

measured, and a training set. Skin detectors were implemented and tested in Adaptive Vision Studio (available at www.adaptive-vision.com). Skin detection effectiveness is measured by: false positive rate (δ_{FP}), which is a percentage of non-skin pixels classified as skin pixels, and false negative rate (δ_{FN}), which is a percentage of skin pixels classified as non-skin. Sum of these rates is treated as detection error used for comparison. In case of statistical models the rates depend on acceptance threshold.

During the experiments several skin detectors were tested: *RGB* detector [10], *HSV* detector [14], a detector in nonlinearly transformed YC_bC_r space [6], composed detector [11], statistical skin model in *RGB*, *HSV* and YC_bC_r and the face-based model developed in the current research. The results are compared in Fig. 4. Performance of statistical models was better than for parametric methods, but the best results were achieved with the face-based model joined with statistical *RGB*. For statistical models acceptance threshold was set to a value, for which sum of the errors was minimal. Choice of a color space for the statistical model do not influence the results significantly. Number of histogram bins per channel is quite relevant, as shown in Fig. 5b for the *RGB* statistical model.

Table 1. False positive error for various false negative rates

Skin detection model	$\delta_{FN} = 5\%$	$\delta_{FN} = 10\%$	$\delta_{FN} = 20\%$
Statistical <i>RGB</i>	23.8%	15.7%	8.8%
Face-based	26.1%	12.6%	5.4%
Face-based with stat. <i>RGB</i>	15.3%	5.1%	1.9%

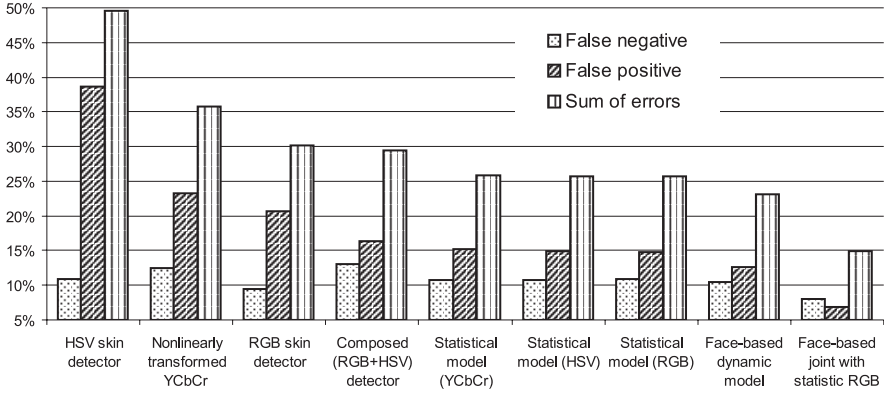


Fig. 4. Comparison of major skin detection methods

Results for *RGB* statistical model, face-based dynamic model and joined model are presented in Fig. 5a in a form of receiver operating characteristic (ROC) curves. Exact values of false positive errors for three false negative rates are presented in Tab. 1.

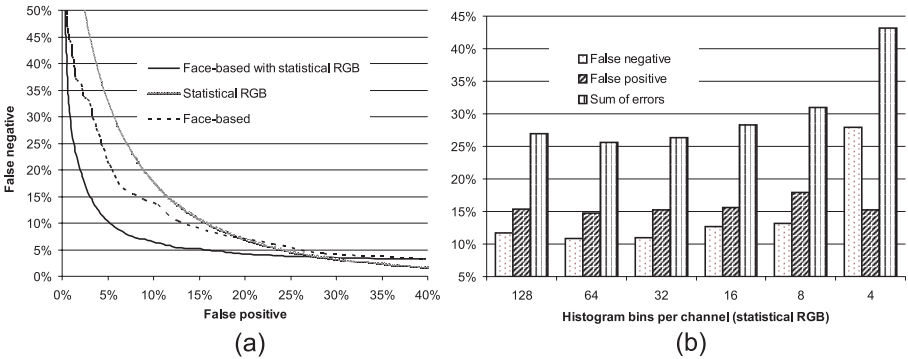


Fig. 5. ROC curves for *RGB* statistical model and face-based models (a). Performance of *RGB* statistical model for various number of histogram bins per channel (b).

6 Conclusions and Future Work

In this paper a concept of dynamic skin detection was presented. Based on face detection performed with Hough transform and the SVM verifier, a statistical skin model is build for every detected face region. After that a general statistical skin model is updated with the face-based model in *RGB* color space. The

developed method was compared with other known skin detectors and it allowed to reduce false positive detection to only 5.1% at 10% false negative rate for the tested skin database.

Future works will be aimed at taking into account not only color information of pixels, but also their neighborhood. It is also planned to detect position of hands in the areas indicated as human skin which will be the next step towards the sign language recognition system.

References

1. Albiol, A., Torres, L., Delp, E.J.: Optimum color spaces for skin detection. In: Proc. of the Int. Conf. on Image Proc., vol. 1, pp. 122–124 (2001)
2. Brand, J., Mason, J.: A Comparative Assessment of Three Approaches to Pixel-level Human Skin-Detection. In: Proc. of the Int. Conf. on Pattern Recognition, vol. 1, pp. 1056–1059 (2000)
3. Cortes, C., Vapnik, V.: Support-Vector Networks. *Machine Learning* 20(3), 273–297 (1995)
4. Filipe, T., Tiago, C., Hamid, S.: Improved Automatic Skin Detection in Color Images. In: VIIIth digital image comp.: techniques & applications, pp. 419–427 (December 2003)
5. Gomez, G., Moralez, E.: Automatic feature construction and a simple rule induction algorithm for skin detection. In: Proc. of the ICML Workshop on Machine Learning in Comp. Vision, pp. 31–38 (2002)
6. Hsu, R.-L., Abdel-Mottaleb, M., Jain, A.K.: Face detection in color images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 24(5), 696–706 (2002)
7. Illingworth, J., Kittler, J.: A survey of the Hough transform. *Comp. Vision Graphics and Image Proc.* 44(1), 87–116 (1988)
8. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. *International Journal of Comp. Vision* 46(1), 81–96 (2002)
9. Kawulok, M.: Application of Support Vector Machines in automatic human face recognition. *Medical Informatics & Technologies* 9, 143–150 (2005)
10. Kovac, J., Peer, P., Solina, F.: Human Skin Colour Clustering for Face Detection. In: Eurocon 2003 Ljubljana, pp. 144–148. Slovenia (2003)
11. Kukharev, G., Nowosielski, A.: Fast and Efficient Algorithm For Face Detection in Colour Images. *Machine Graphics & Vision* 13(4), 377–399 (2004)
12. Phung, S.L.: ECU Face Detection Database, University of Wollongong (2002)
13. Terrillon, J.-C., David, M., Akamatsu, S.: Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments. In: Proc. of the 3rd International Conf. on Automatic Face and Gesture Recognition, Nara, Japan, pp. 112–117 (1998)
14. Tsekeridou, S., Pitas, I.: Facial feature extraction in frontal views using biometric analogies. In: Proc. of EUSIPCO 1998, vol. 1, pp. 315–318 (September 1998)
15. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. Tech. Report CARTR-948, Center for Automation Research, University of Maryland, College Park (2000)