

Texture Analysis for Skin Probability Maps Refinement

Michał Kawulok*

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

Abstract. In this paper a new method for skin regions detection and segmentation is proposed. To improve the conventional color-based skin models, skin probability maps are subject to texture analysis using discriminative statistical features. Although the texture was utilized for skin detection in some of the existing methods, the main contribution of the work reported here is that the probability maps rather than the original color images are processed. The method has been validated in a series of experiments using two data sets. The obtained results are reported in the paper, and they confirm that the method is competitive.

1 Introduction

Skin region detection and segmentation has a wide range of potential applications, and it has received considerable attention from the computer vision community. The general aim is to classify every individual pixel in a given digital image as belonging to skin or not. In particular, skin detection is an important source of information for locating faces and hands for human-computer interaction purposes, being the first step of the image understanding chain. Segmenting skin regions can also be used for indexing images in multimedia databases, for content-based filtering and parental control applications.

There have been many skin detection methods proposed which are based on modeling skin color in various color spaces. Such models are later used to determine the probability that a given pixel presents human skin. In this way, the input image is converted into a skin presence probability map. Depending on a particular approach, such a map may have continuous or binary values. Skin color models are usually quite sensitive to illumination variations, and alternative sources of information can be helpful to stabilize the result.

In the work reported here we propose and investigate analysis of *discriminative textural features* (DTF) performed in the probability map obtained using a statistical approach [1]. We analyze the texture based on simple statistics computed in kernels of various size. The statistical features are later subject to *linear discriminant analysis* (LDA). This generates the LDA feature space, in which we

* This work has been supported by the Polish Ministry of Science and Higher Education under research grant no. IP2011 023071 from the Science Budget 2012–2013.

measure the similarity between the pixels with regards to their skin-likeness. In this way we managed to reduce the segmentation errors substantially, which was confirmed experimentally.

The paper is organized as follows. In Section 2 an overview of existing approaches to skin segmentation is given, and statistical skin segmentation that is the basis for the research reported here is described in Section 3. Our contribution is explained in Section 4 and experimental validation results are shown and discussed in Section 5. The paper is concluded and directions of future work are presented in Section 6.

2 Related Work

Existing skin segmentation techniques take advantage from the observation that skin-tone color has common properties which can be defined in various color spaces. In general, skin color detectors are based on parametric or statistical skin modeling. An interesting, thorough survey which compares various color-based skin detection routines was presented in 2007 by Kakumanu et al. [2].

Parametric skin models are based on fixed decision rules defined empirically in various color spaces after skin-tone distribution analysis. These rules are applied after color normalization to determine if a pixel color value belongs to the skin. Kovac et al. [3] proposed a model defined in RGB color space. Skin-tone color was also modeled in HSV by Tsekeridou et al. [4]. An approach proposed by Hsu et al. [5] takes advantage of common skin color properties in nonlinearly transformed YC_bC_r color space, in which elliptical skin color model is defined. Some techniques operate in multiple color spaces to increase the stability, for example a composed skin detector [6] defined in RGB and YC_bC_r color spaces. Recently, Cheddad et al. proposed to reduce the RGB color space to a single dimension, in which the decision rules are defined [7].

Statistical modeling is based on analysis of skin pixel values distribution for a training set of images, in which skin and non-skin areas are already identified and marked. This creates a global model of skin color, which allows determining the probability that a given pixel value belongs to the skin class. Skin color can be modeled using a number of techniques, including Bayes classifier [1], Gaussian mixture model [8] or random forests [9]. The first method [1] was used in the research reported here to generate the probability maps and it is given more attention in Section 3.

There are a number of adaptive models that improve the segmentation accuracy. Lee et al. proposed to extract lighting features from every analyzed image to adjust the skin detector [10]. Phung et al. introduced a method for adapting the segmentation threshold in the probability map [11], and this approach was later extended by Zhang et al. [12]. We also investigated adaptive approaches during our earlier works. First, we proposed to adjust the global skin model to every face detected in the image [13], and then we managed to decrease false positives using skin blobs analysis [14].

False positive errors can be reduced by analyzing neighborhood of pixels classified as skin [15]. Also, the features helpful for skin segmentation can be extracted

using texture analysis performed in grayscale [16, 17] or color domain [18]. However, in these methods the skin probability maps were not analyzed with regards to their textural features.

When skin segmentation is performed in video sequences, the system may take advantage of dynamic information. Sigal et al. used Markov models to predict illumination changes in subsequent video frames to adjust the skin color model [19]. Furthermore, background extraction techniques and motion detectors may be used to find potential locations of skin pixels.

3 Statistical Skin Modeling

In the research reported here, statistical skin modeling based on Bayes classifier [1] was used to obtain the skin probability maps that were later analyzed and refined. The method requires training to determine skin probability lookup table. At first, based on an annotated training set, histograms for skin (C_S) and non-skin (C_{NS}) classes are built. For every color value (v) the probability of observing it in each class equals:

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

where $C_x(v)$ is a number of v -color pixels in the x -th class and N_x is a 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, we determined experimentally that detection is more effective if the number of bins is reduced to 64 per channel. This was studied in our earlier work [14] and it is also justified in [1]. To detect skin with low false positive error, not only is it important that a pixel value appears frequently as skin, but also its density in the non-skin class should be low. Hence, the probability that a given pixel value belongs to the skin class is obtained using 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 large enough, the probabilities for all possible color values can be computed to create skin color probability lookup table.

After obtaining the probability lookup table, every pixel value can be transformed into the probability that this pixel belongs to the skin class. Skin regions can be later extracted based on a fixed acceptance threshold or more advanced thresholding and blob detection techniques.

4 Texture-Based Probability Map Processing

Distribution of skin color has been effectively modeled in various color spaces and chrominance combined with the luminance component is considered as the

most distinctive skin feature. However, its discriminating capability is limited, because skin color is not exclusive for skin regions and other objects can also have similar color to human skin. Hence, color-based skin models are quite effective for detecting skin pixels, however they fail when the background objects are of skin-like color. In order to decrease false positives, the model should be extended to use other distinctive features extracted in pixels' neighborhood, instead of classifying every individual pixel independently.

We explored how to enhance the conventional color model using textural features. Although human skin is characterized by distinctive pattern, it can be observed only in high-quality images of high resolution. Otherwise, in images of medium or low quality, skin regions are rather plain, without any remarkable textural properties. On the other hand, false positives often present relatively small color variations which are well manifested in the skin color probability map. This is illustrated in Fig. 1, in which an original image (a), and corresponding skin probability map (b) are presented (darker shade in the map indicates higher probability). The selected region, characterized by a smooth texture with small chrominance variations, is magnified in the Figure. It may be noticed that although the values in the probability map are high within the region, they form a distinctive pattern, which cannot be observed for real skin. Such phenomenon appears in case of many images, which leads to a conclusion that textural features extracted from the probability map may be helpful for skin segmentation. During our experiments we found it much more beneficial to extract these features from the probability maps rather than from the original images.



Fig. 1. Input image (a) and obtained skin probability map (b)

In order to determine which textural features discriminate well between skin and non-skin regions, we adapted a framework which we developed for image colorization purposes [20]. First, we compute simple image statistics in several kernels of different dimension. This forms *basic image features* (BIF) which are subsequently subject to linear discriminant analysis to obtain discriminative textural features. The details of the proposed feature extraction technique are presented later in this Section.

4.1 Linear Discriminant Analysis

Linear discriminant analysis [21] is a supervised statistical feature extraction method frequently used in machine learning. It finds a subspace defined by the most discriminative directions within a given training set of M -dimensional vectors classified into K classes. The analysis is performed first by computing two covariance matrices: within-class scatter matrix $S_W = \sum_{i=1}^K \sum_{\mathbf{u}_k \in K_i} (\mathbf{u}_k - \boldsymbol{\mu}_i)(\mathbf{u}_k - \boldsymbol{\mu}_i)^T$ and between-class scatter matrix $S_B = \sum_{i=1}^K (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T$, where $\boldsymbol{\mu}$ is a mean vector of the training set and $\boldsymbol{\mu}_i$ is a mean vector of the i -th class (termed K_i). Subsequently, the matrix $S = S_W^{-1} S_B$ is subjected to the eigen decomposition $S = \Phi \Lambda \Phi^T$, where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_M)$ is the matrix with the ordered eigenvalues along the matrix diagonal and $\Phi = [\mathbf{v}_1 | \dots | \mathbf{v}_M]$ is the matrix with the correspondingly ordered eigenvectors as columns. The eigenvectors form the orthogonal basis of the feature space.

Originally, the feature space has M dimensions, but only those associated with the highest eigenvalues have strong discriminative power, while the remaining can be rejected. In this way the dimensionality is reduced from M to m , where $m < M$.

After having built the m -dimensional feature space, the feature vectors are obtained by projecting the original vectors \mathbf{u} onto the feature space: $\boldsymbol{\nu} = \Phi^T \mathbf{u}$. The similarity between the feature vectors is computed based on their Euclidean distance in the feature space.

4.2 Discriminative Textural Features

In order to determine the discriminative textural features, first the basic image features are calculated from every pixel in the probability map. BIF are composed of: a) *mean* and b) *median* value, c) *standard deviation*, d) *minimal* and e) *maximal* value, computed in multiple kernels of different size, and f) the *difference between maximum and minimum* values in each kernel. These six features are obtained for four kernel sizes ranging from 5×5 to 11×11 . Hence, every pixel x is described with an M -dimensional basic feature vector u_x ($M = 24$ in the presented case).

Skin and non-skin pixels are transformed into two classes of BIF vectors which are subsequently subject to LDA in order to increase their discriminative capability. This training process, illustrated in Fig. 2a, creates a discriminative feature space that is later used to generate skin probability maps. First, for a given training set of images and associated skin masks, the probability lookup table is obtained using statistical skin modeling [1]. After that, all the images from the training set are transformed into probability maps, from which the basic image features are extracted. Finally, based on the basic feature vectors, labeled using the ground-truth skin masks, LDA projection matrix is computed. Based on the eigenvalues, it was observed that the discriminating power was cumulated in the first three components, hence the dimensionality of the DTF space was reduced to $m = 3$.

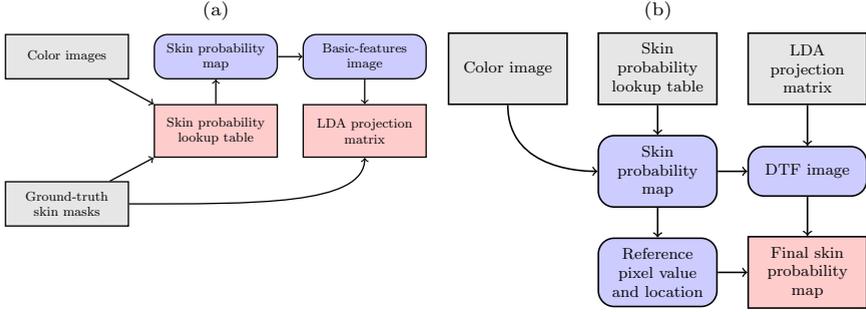


Fig. 2. Flowchart of the training (a) and detection (b) process

4.3 Skin Detection Using DTF Space

Skin detection in the proposed method is implemented in two general stages, namely: 1) probability map generation using statistical model, and 2) probability map refinement in the DTF space. A flowchart for this process is presented in Fig. 2b and it is also illustrated using an example in Fig. 3. First of all, an original image (Fig. 3a) is converted into the skin probability map (Fig. 3b) using the lookup table obtained after training the statistical model. Subsequently, every pixel from the probability map is projected onto the DTF space based on the LDA projection matrix.

The probability maps are refined based on the distance of every pixel from a reference pixel r , computed in the DTF space. The reference pixel is determined as a pixel of maximal value in the probability map subjected to a minimum filter of a large 15×15 kernel (Fig. 3c). The filtering is applied, because the reference pixel should not be an isolated skin spot, as the surrounding non-skin pixels could disturb the textural features. Therefore, the filter kernel is larger than the largest kernel used for BIF extraction. Finally, the skin probability map is obtained based on the distances from the reference pixel in the DTF space, and the maximal probability in the refined map is equal to the reference pixel probability before the filtering.

We investigated a possibility of using support vector machines (SVM) for classifying the vectors in the DTF space to determine whether they belong to the skin. This required an additional step during training (i.e. the SVM was trained using labeled pixels from the train set, represented in the DTF space), but it also introduced substantial burden on the skin detector, as every individual pixel had to be classified with the SVM. We found it equally effective and much more efficient to use the DTF space only to determine similarity between the pixels with regards to their “skinniness”, as explained earlier.

5 Experimental Validation

The experiments were carried out for two datasets: 4000 images from the ECU database [11], and 899 hand images (termed HGR) which we registered for

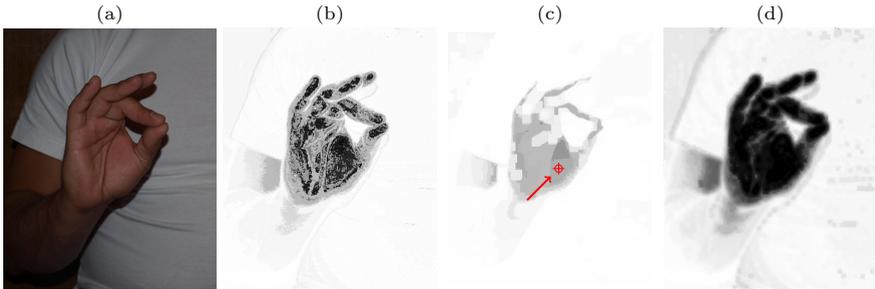


Fig. 3. Probability maps refinement: a) an original image, b) the probability map, c) a reference pixel in the filtered map, and d) the refined probability map

gesture recognition purposes. The images from both datasets are associated with ground-truth skin binary masks indicating skin regions. The ECU images were acquired in uncontrolled lighting conditions, and skin-color objects often appear in the background, which makes the skin regions difficult for segmentation. The HGR data set contains images of gestures presented by 12 different people. The data were acquired in controlled conditions, but in some cases the subjects were dressed in clothes of skin-like color. The proposed algorithm was compared with the statistical method [1] and several parametric skin color models [3–7].

For methods which require learning the images were randomly split into train and test sets, and the results were obtained using 10-fold cross validation. Skin segmentation performance was assessed based on two errors, namely: a) false positive rate (δ_{FP}), obtained as a percentage of background pixels classified as skin, and b) false negative rate (δ_{FN}), i.e. a percentage of skin pixels misclassified as background. False positives and false negatives both depend from the acceptance threshold, and their mutual relation is presented using *receiver operating characteristics* (ROC).

The method was trained using 200 images from the ECU database and validated with the remaining images from the ECU and HGR data sets. The HGR data set was not used for training as it contains many similar images of the same individuals, acquired in uniform conditions. ROC curves are shown in Fig. 4, false positive rates obtained for various false negative thresholds are given in Tab. 1, and the scores for the parametric methods are presented in Tab. 2. The latter do not depend on the acceptance threshold, so they appear as points in Fig. 4. It may be noticed that the proposed method reduces the false positives significantly for fixed false negative rates, and it is also competitive compared with the parametric methods.

Several examples of error reduction using the proposed approach are demonstrated in Fig. 5 for images from the ECU database and in Fig. 6 for images from our HGR data set. An original image (I.) is transformed into a probability map (II.), which is thresholded to obtain the skin regions. Images before (III.) and after (IV.) the proposed DTF-based refinement of the probability maps are presented in the Figures. Correctly detected skin regions (i.e. true positives) are

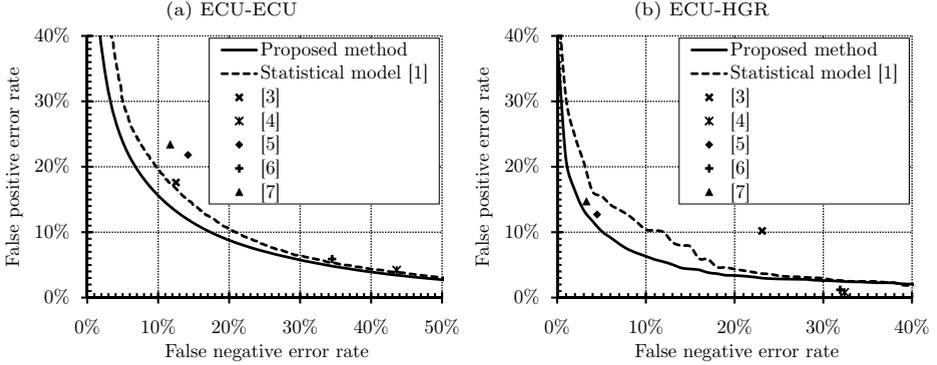


Fig. 4. ROC curves obtained for the ECU and HGR data sets

Table 1. False positive error rate for different false negative thresholds

Train set	Test set	δ_{FN}	Statistical model [1]	Proposed method
ECU	ECU	5%	$29.94 \pm 1.96\%$	$23.44 \pm 0.66\%$
		10%	$19.59 \pm 0.68\%$	$15.15 \pm 0.71\%$
		15%	$13.95 \pm 0.46\%$	$11.26 \pm 0.83\%$
		20%	$10.45 \pm 0.33\%$	$8.78 \pm 0.68\%$
ECU	HGR	5%	$15.87 \pm 1.9\%$	$10.86 \pm 2.97\%$
		10%	$11.43 \pm 3.89\%$	$6.56 \pm 2.15\%$
		15%	$8.53 \pm 2.15\%$	$4.56 \pm 1.82\%$
		20%	$5.4 \pm 1.11\%$	$3.39 \pm 1.44\%$

Table 2. Detection errors obtained using parametric models

Test set	Error type	[3]	[4]	[5]	[6]	[7]
ECU	δ_{FN}	12.5%	43.6%	14.2%	34.5%	11.7%
	δ_{FP}	17.6%	4.2%	21.8%	5.9%	23.4%
HGR	δ_{FN}	23.1%	32.4%	4.5%	31.9%	3.3%
	δ_{FP}	10.2%	0.8%	12.7%	1.2%	14.7%

annotated with green borders, false positives are presented in a red shade, and false negatives – in a blue shade. True negatives are shown with faded color. It can be noticed from the Figures that the proposed method reduces both false negatives and false positives for the presented images. Due to the large kernel size used for extracting the basic features, the edges of the images are not processed, which explains why in some cases these areas remain unimproved. In the case of the HGR images, the segmentation often delivered high false negatives using a global threshold as in Fig. 6a (III.). On the other hand, a lower value of the acceptance threshold resulted in high false positives in other images as in Fig. 6b (III.). However, for the both cases the result is much better after the refinement (IV.).

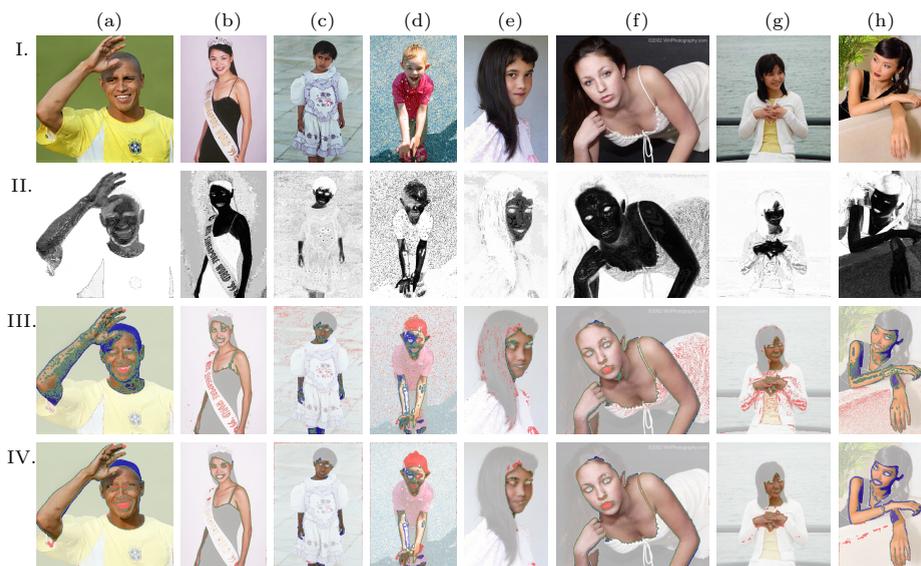


Fig. 5. Examples of results obtained for ECU images (from top: I. original image, II. probability map, result before (III.) and after (IV.) probability map refinement)

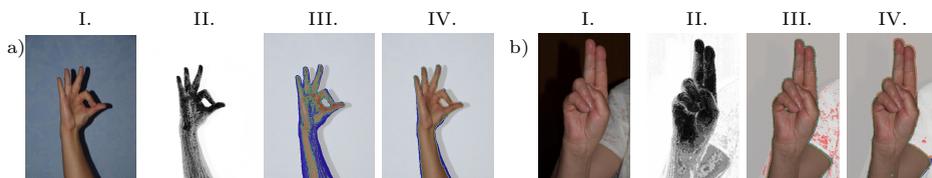


Fig. 6. Examples of results obtained for HGR images (I. original image, II. probability map, result before (III.) and after (IV.) probability map refinement)

6 Conclusions and Future Work

In this paper a new method for improving skin segmentation has been proposed. This aim was achieved by exploiting discriminative textural features extracted from the skin probability maps. The presented experimental results demonstrate that the errors can be substantially reduced using the proposed approach.

In addition to the error reduction presented in the paper, the refined probability maps may constitute a better input for skin blobs analysis, which will be one of the directions for our future work. We will also explore a possibility of segmenting skin regions using the distance transform in the DTF space from a set of reference points. Furthermore, only simple image statistics were used as the basic textural features here, and we plan to investigate application of alternative features, especially those that represent local frequency.

References

1. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. *Int. J. of Comp. Vision* 46, 81–96 (2002)
2. Kakumanu, P., Makrogiannis, S., Bourbakis, N.G.: A survey of skin-color modeling and detection methods. *Pattern Recognition* 40(3), 1106–1122 (2007)
3. Kovac, J., Peer, P., Solina, F.: Human skin color clustering for face detection. In: *EUROCON 2003. Computer as a Tool*, vol. 2, pp. 144–148 (September 2003)
4. Tsekeridou, S., Pitas, I.: Facial feature extraction in frontal views using biometric analogies. In: *Proc. of EUSIPCO 1998*, pp. 315–318 (1998)
5. Hsu, R.-L., Abdel-Mottaleb, M., Jain, A.K.: Face detection in color images. *IEEE Trans. Pattern Anal. and Machine Intelligence* 24(5), 696–706 (2002)
6. Kukharev, G., Nowosielski, A.: Fast and efficient algorithm for face detection in colour images. *Machine Graphics and Vision* 13, 377–399 (2004)
7. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: A skin tone detection algorithm for an adaptive approach to steganography. *Signal Processing* 89(12), 2465–2478 (2009)
8. Greenspan, H., Goldberger, J., Eshet, I.: Mixture model for face-color modeling and segmentation. *Pattern Recogn. Lett.* 22, 1525–1536 (2001)
9. Khan, R., Hanbury, A., Stöttinger, J.: Skin detection: A random forest approach. In: *Proc. 17th IEEE Int. Image Processing (ICIP) Conf.*, pp. 4613–4616 (2010)
10. Lee, J.S., Kuo, Y.M., Chung, P.C., Chen, E.L.: Naked image detection based on adaptive and extensible skin color model. *Pattern Recognition* 40, 2261–2270 (2007)
11. Phung, S.L., Chai, D., Bouzerdoum, A.: Adaptive skin segmentation in color images. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 353–356 (2003)
12. Zhang, M.-J., Wang, W.-Q., Zheng, Q.-F., Gao, W.: Skin-color detection based on adaptive thresholds. In: *Proceedings of the Third International Conference on Image and Graphics, ICIG 2004*, pp. 250–253. IEEE Computer Society, Washington, DC (2004)
13. Kawulok, M.: Dynamic Skin Detection in Color Images for Sign Language Recognition. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D. (eds.) *ICISP 2008. LNCS*, vol. 5099, pp. 112–119. Springer, Heidelberg (2008)
14. Kawulok, M.: Energy-based blob analysis for improving precision of skin segmentation. *Multimedia Tools and Applications* 49(3), 463–481 (2010)
15. del Solar, J.-R., Verschae, R.: Skin detection using neighborhood information. In: *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 463–468 (2004)
16. Wang, X., Zhang, X., Yao, J.: Skin color detection under complex background. In: *Proc. International Conference on Mechatronic Science, Electric Engineering and Computer* (2011)
17. Abin, A.A., Fotouhi, M., Kasaei, S.: A new dynamic cellular learning automata-based skin detector. *Multimedia Syst.* 15(5), 309–323 (2009)
18. Conci, A., Nunes, E., Pantrigo, J.J., Sánchez, Á.: Comparing color and texture-based algorithms for human skin detection. In: *ICEIS* (5), pp. 166–173 (2008)
19. Sigal, L., Sclaroff, S., Athitsos, V.: Skin color-based video segmentation under time-varying illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 862–877 (2003)
20. Kawulok, M., Smolka, B.: Texture-adaptive image colorization framework. In: *EURASIP Journal on Advances in Signal Processing* 2011(99) (2011)
21. Seber, G.: *Multivariate Observations*. Wiley, New York (1984)