Self-Adaptive Skin Segmentation in Color Images

Michal Kawulok[1], Jolanta Kawulok[1,*],
Jakub Nalepa[1,2], and Bogdan Smolka[1,**]

[1] Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology, Gliwice, Poland
{michal.kawulok,jolanta.kawulok,jakub.nalepa,bogdan.smolka}@polsl.pl
[2] Future Processing, Gliwice, Poland

Abstract. In this paper, we present a new method for skin detection and segmentation, relying on spatial analysis of skin-tone pixels. Our contribution lies in introducing self-adaptive seeds, from which the skin probability is propagated using the distance transform. The seeds are determined from a local skin color model that is learned on-line from a presented image, without requiring any additional information. This is in contrast to the existing methods that need a skin sample for the adaptation, e.g., acquired using a face detector. In our experimental study, we obtained F-score of over 0.85 for the ECU benchmark, and this is highly competitive compared with several state-of-the-art methods.

Keywords: skin detection, skin segmentation, skin color, adaptive skin modeling, spatial analysis, distance transform, gesture recognition.

1 Introduction

Detection and segmentation of human skin regions is a challenging computer vision and pattern recognition task, which has attracted a number of contributions over the years. Potential applications include gesture recognition, objectionable content filtering, image retrieval, image coding, and more [1, 14].

Existing approaches are based on modeling the skin color in various color spaces in order to discriminate between skin and non-skin pixels based on their position in the color space [8]. Skin color models can be defined as a set of rules and thresholds [2,3,13], or they may be generated using machine learning. Given a sufficiently large training set, the Bayesian classifier is effective here [7], while for smaller sets Gaussian mixture models (GMMs) are often used [5]. In general, these methods, as well as other machine learners, transform an input color image into a *skin probability map* (SPM) that is binarized to extract the skin regions.

Although the chrominance is considered as the most discriminative skin presence feature, the efficacy of pixel-wise classification schemes is limited due to low specificity and high variance of the skin color. This has been addressed in several ways. A skin color model can be dynamically adapted to each presented image based on a skin sample acquired using face [11,12,19] or hand [1] detectors. From a detected skin sample, a local skin model is learned, combined with the global one and applied to the image. In [19], the face-based adaptation is performed in the error signal (ES) space [2]. Alternatively, the lighting conditions are analyzed to select the most suitable skin model [14], or to set its parameters [18].

Adaptive skin color modeling increases the specificity, however in most cases the skin cannot be completely separated from the background in a color space [20]. The errors can also be reduced by analyzing the neighborhood, either relying on the texture [17] or using spatial analysis. The latter takes advantage of the fact that skin pixels form consistent regions [9, 16]. First, skin seeds are determined using a high-probability threshold, and then the skin probability (termed *skinness*, P_S) is propagated from them. This improves the detection, but the outcome is sensitive to the seed extraction procedure. Basically, a seed should appear in every skin region, and errors committed here cannot be fixed during propagation.

There have been some attempts to combine different improvement techniques. Color, texture and spatial analysis were used in a system proposed in [6]. We proposed to extract textural features from SPMs and exploit them to compute the weights for the distance transform (DT) [10]. Two other methods combine face-based model adaptation with the spatial analysis. In [12], a local model learned from a facial region is used to obtain the foreground weights for the graph-cut image segmentation. The background weights are obtained using a global model. In [11], we applied a local model to the image to determine the seeds for the DT, whose weights are computed based on the global model [9].

In the work reported here, we introduce a new method for generating the skinness propagation seeds. First, we analyze SPMs, obtained using a global model, to determine the *initial skin seeds* (ISSs). Next, the ISSs are enlarged using the DT to include more skin pixels for training a local skin model. Subsequently, this model is applied to obtain the final adapted seeds, and the skinness is propagated over the entire image. Overall, our contribution consists in extracting skin samples for the adaptation directly from SPMs without relying on face or hand detectors. Results of our experimental study indicate that the achieved detection accuracy is comparable with that obtained using face-based methods, and definitely better than in case of other state-of-the-art techniques.

In Section 2, the baseline techniques exploited in our study are outlined. The proposed skin detection method is described in Section 3, the experimental study is reported in Section 4, and the paper is concluded in Section 5.

2   Baseline Methods

In this section, we briefly present our earlier strategy to combine model adaptation with spatial analysis [11]. First, a local skin color model is learned from

the facial region pixels, using a single multivariate Gaussian. Afterwards, we compute the skin probability based on the pixel color c in RGB:

$$P_S(c) = \exp\left[-0.5(c - \overline{c})^T \Sigma^{-1}(c - \overline{c})\right] / \sqrt{(2\pi)^3 |\Sigma|}, \qquad (1)$$

where $\overline{c}$ is the mean color value and Σ is the color covariance matrix of the skin pixels. The local model detects skin with high precision, but low recall, hence it serves to determine the seeds, from which the skinness is propagated in the image using the Dijkstra's algorithm [9]. We determine the skinness propagation paths so as to minimize the sum of local distances between the neighboring pixels (p_i and p_{i+1}) along the path: $\Gamma = \sum_{i=0}^{l-1} \gamma(p_i \to p_{i+1})$. The skin probability is determined based on Γ ($P_S = 1$ for the seeds, where $\Gamma = 0$, and $P_S = 0$ for Γ_{max}). The local costs from pixel x to y are computed taking into account two components: (i) obtained from the color image (γ_c), and (ii) from the SPM (γ_p): $\gamma(x \to y) = \gamma_c(x, y) \cdot [1 + \gamma_p(x \to y)]$. Here, we use two different costs γ_c, discussed later in Section 3, and we adopt the SPM cost introduced in [9]:

$$\gamma_p(x \to y) = \begin{cases} 1 - P_S(y) \text{ for } P_S(y) > P_\beta \\ \infty \quad \text{ for } P_S(y) \leq P_\beta \end{cases}, \qquad (2)$$

where P_β is the propagation threshold.

3   Skin Segmentation Using Self-adaptive Skin Seeds

A flowchart of our method is presented in Fig. 1, and the process is illustrated using two examples in Fig. 2. First of all, an input image (Fig. 2a) is converted into an SPM (Fig. 2b) using a global skin color model (we used the Bayesian classifier). The obtained SPM is analyzed to determine the ISSs, i.e., small skin patches (black pixels inside the gray regions in Fig. 2c). The general goal here is to obtain a skin sample at the smallest possible false positive rate, and the exact procedure is described later in Section 3.1. Subsequently, the ISSs are expanded using the DT to include more skin pixels (gray regions in Fig. 2c).

From the enlarged seeds, a local GMM is trained on-line and applied in order to determine the final seeds for the propagation (black regions in Fig. 2d). Here, the goal is to find at least a single seed in every skin region, while keeping the false positives low (see Section 3.2). From the final seeds, the skinness is propagated over the image to obtain the final SPM (Fig. 2e) and the skin regions are extracted (Fig. 2f, where red indicates false positives and blue – false negatives).

In Fig. 2g, we present the segmentation result obtained from the global SPM. For the image in the upper row, the adaptation reduced false positives (the background objects have skin-like color), while in the bottom row, false negatives were decreased (the skin does not have a typical color here due to a flash light).

3.1   Initial Skin Seeds Extraction

This step consists in finding skin samples, which in alternative methods [11, 19] are acquired from a facial region. Our intention is to perform the adaptation
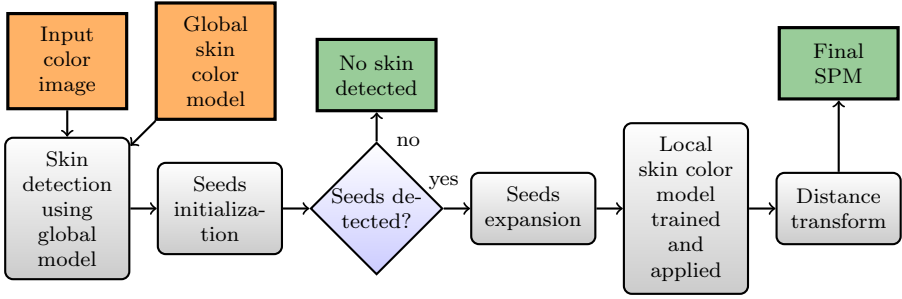
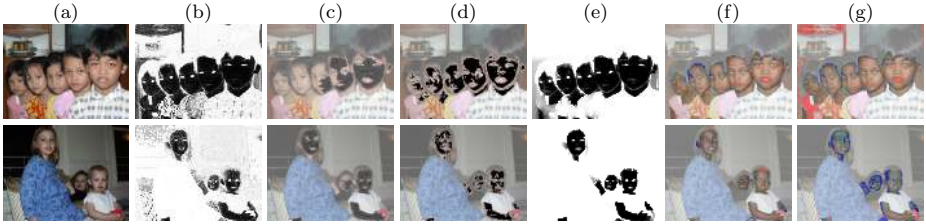**Fig. 1.** Flowchart of the proposed skin segmentation process

Fig. 2. Subsequent steps of the process: input color image (a), global SPM (b), ISSs before (black) and after (gray) the DT-based expansion (c), seeds after adaptation (d), final SPM (e), and segmentation result from the adapted (f) and global SPM (g)

without the necessity of engaging face detectors, and we this can effectively be done relying on the SPM obtained using a global model. The procedure is presented in Alg. 1. First, we compute the integrated histogram of the SPM to find a threshold value (t_s) that selects $\mathcal{R}_s = 5\%$ out of all the pixels with the highest probability. In order to avoid adapting to images without any skin pixels, we assume that t_s cannot be less than $\mathcal{T}_{min} = 0.6$. Otherwise, the original SPM is used. We have observed that false positives are scattered after the binarization, while true positive pixels are organized in spatially consistent groups. Hence, we use only 10% of the largest blobs as the ISSs, and the seeds are subject to the morphological skeletonization to further reduce the false positive rate.

3.2 Seeds Expansion and Adaptation

The ISSs indicate the skin regions with high precision, but they do not appear in every skin blob (see Fig. 2c). Hence, a local skin model is created to propagate the skinness not only in the spatial domain, but also across the color space. The ISSs contain too few pixels to build a local model that would detect the seeds in every skin region, hence the seeds are first expanded using the DT, outlined earlier in Section 2. To minimize the number of adjoined non-skin pixels during the expansion, we use a restrictive local cost in 4-neighborhood:

$$\gamma_c^{(YH)}(x,y) = \max\left(|Y(x) - Y(y)|, |H(x) - H(y)|\right), \tag{3}$$

Algorithm 1. Seeds initialization

Require: $\mathbf{P_S}$	▷ The SPM obtained using a global model
1: $H_I \leftarrow$ IntegratedHistogram($\mathbf{P_S}$);	▷ Compute integrated histogram of the SPM
2: $t_s \leftarrow$ GetHighest(H_I, $\mathcal{R}_s$);	▷ $\mathcal{R}_s$ (%) pixels have skin probability over t_s
3: **if** $t_s \geq \mathcal{T}_{min}$ **then**	
4: $\mathbf{P_S} \leftarrow$ Binarize($\mathbf{P_S}$, t_s);	▷ Binarize the SPM with a threshold t_s
5: $\mathbf{B} \leftarrow$ FindBlobs($\mathbf{P_S}$);	▷ Find a set of blobs $\mathbf{B}$ in the ISSs image
6: $\mathbf{B} \leftarrow$ FilterBlobSize($\mathbf{B}$);	▷ Select 10% of the largest blobs
7: $\mathbf{P_S} \leftarrow$ Render($\mathbf{B}$);	▷ $\mathbf{P_S}$ gets a mask of the filtered blobs
8: $\mathbf{P_S} \leftarrow$ MorphologicalSkeleton($\mathbf{P_S}$);	▷ Now $\mathbf{P_S}$ presents a mask of the ISSs
9: **end if**	

where $H(\cdot)$ is the hue and $Y(\cdot)$ is the luminance. Also, we ignore the pixels, whose total cost is above $\mathcal{L}_T = 3\overline{\gamma_c}$, where $\overline{\gamma_c}$ is the average local cost in the image.

The enlarged seeds are used to model the skin distribution with a single multivariate Gaussian, and the model is applied to the entire image to extract the final seeds. In order to avoid propagation from isolated pixels, the seeds are subject to the erosion using 5×5 kernel. From these seeds, the skinness is propagated using less restrictive local cost computed in the RGB color space:

$$\gamma_c^{(RGB)}(x, y) = (|R(x) - R(y)| + |G(x) - G(y)| + |B(x) - B(y)|). \qquad (4)$$

After propagation, the distance map is normalized to obtain the final SPM.

4   Experimental Validation

In our experiments, we used the ECU benchmark set [15]. Out of 4000 images with annotated ground-truth skin regions, 2000 were used to train the Bayesian classifier (there were $1.0 \cdot 10^8$ skin and $4.7 \cdot 10^8$ non-skin pixels in the training set), and the remaining 2000 images formed a test set. We used the same setup[1] as in our earlier work [11], and we set $P_\beta = 0.3$ for spatial analysis as justified in [10]. To assess the skin detection performance, we used the F-score, precision, recall, and false positive rate (δ_{fp} – the percentage of background pixels classified as skin). Also, we verified whether the proposed method adapts to the images which do not present the skin at all. To do so, we removed the skin pixels from the tested images, and we report the false positive rate for these images (δ_{fp}^{NS}). Naturally, $\delta_{fp}^{NS} = \delta_{fp}$ in case of pixel-wise schemes. The experiments were run using a computer equipped with an Intel Core i7-3740QM 2.7 GHz (16 GB RAM) processor.

The obtained results are reported in Tab. 1, and most relevant precision-recall curves [4] are shown in Fig. 3. In the table, the acceptance thresholds used to binarize the SPMs were set so as to maximize the F-score. We categorize the methods into several groups, namely: (1) the global pixel-wise schemes, (2) texture and spatial analysis, (3) adaptive methods, and (4) the proposed

---

Table 1. Skin detection scores obtained using various methods

Group	Method	F-score	Precision	Recall	δ_{fp}	δ_{fp}^{NS}
(1)	Global Bayesian classifier [7]	0.7772	73.15%	82.89%	9.13%	9.13%
	Global model in the ES space [2]	0.7434	68.07%	81.88%	11.79%	11.79%
	Global rule-based model in RGB [3]	0.6869	55.29%	91.61%	23.11%	23.11%
(2)	Discriminative skin-presence features [10]	0.8305	78.09%	88.69%	5.95%	5.24%
	DT from high-probability seeds [9]	0.8177	75.79%	88.78%	9.61%	10.06%
	Color, texture and spatial information [6]	0.7894	76.34%	81.73%	8.43%	9.01%
(3)	Face-based adaptation in the ES space [19]	0.7672	69.67%	85.34%	14.13%	11.79%
	DT from face-based adaptive seeds [11]	0.8661	82.70%	90.92%	6.53%	9.13%
(4)	Proposed method	0.8543	82.26%	88.85%	5.57%	10.38%
	ISSs	0.9247*	90.99%	94.01%*	—	—
	Enlarged ISSs	0.9345*	91.62%	95.36%*	—	—
	Final seeds	0.9548*	91.91%	99.34%*	—	—

* potential recall, assuming correct propagation from the seeds.
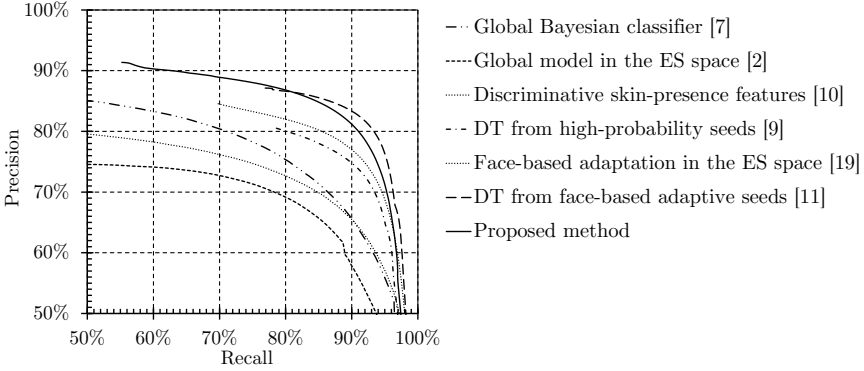
Fig. 3. Precision-recall curves obtained for the ECU data set using various methods

approach. It may be observed that our method outperforms the non-adaptive schemes from groups (1) and (2), as well as the face-based adaptation in the ES space. Although the obtained scores are slightly worse compared with the DT from the face-based adaptive seeds [11], our method does not use any additional information delivered by a face detector. This is an obvious advantage in case of images which do not present a face (e.g., in hand gesture recognition).

It can be seen from Tab. 1 (δ_{fp}^{NS}) that in some cases our method adapts to the images without any skin regions, resulting in some false positives. This is the most important potential pitfall in our approach, however as δ_{fp}^{NS} is only 1.25% larger than using the Bayesian classifier, this is not a significant drawback. In the table, we also report the scores obtained within the seeds. Here, we show a potential recall, considered as an upper bound for the spatial analysis (if a ground-truth skin blob contains a seed inside, then it is regarded as correctly detected). Due to small size of the seeds, δ_{fp} is close to zero, and we do not quote it here. It is worth noting that for subsequent seeds in the processing chain, the scores are improved. This justifies the proposed seed extraction procedure.

Our algorithm processed 512×512 images at 2 fps. However, the adaptation may be carried out only for the first video frame, and then the detector's speed increases to 4 fps, which is similar to the speeds reported in [9, 10].

Fig. 4. Examples of skin detection outcome (red: false positives, blue: false negatives)

In Fig. 4, we present some examples of skin segmentation using various methods grouped as in Tab. 1. In general, the proposed method delivers comparable results to those obtained using face-based adaptation [11]. An interesting example is shown in the second row from the top. Here, both face-based adaptation schemes [11,19] treated the dark shade from the glasses as skin, and as a result, the dark background was also classified as skin. Such problems appear, if the face is occluded or if a face detector does not indicate the facial region precisely. It is worth noting that the introduced method deals well with such cases.

5   Conclusions and Future Work

In this paper, we introduced a new self-adaptive method for detecting skin seeds, and combine it with spatial analysis using the DT. The most significant contribution consists in proposing an adaptation scheme that does not require any explicit skin sample, which makes it independent from face detectors. As a consequence, the method is applicable to images that do not present a face, and may be particularly useful for hand gesture recognition purposes.

Our ongoing work is aimed at incorporating the textural features into the proposed adaptation scheme. Furthermore, our research plans include improving the image-level skin detection so as to reduce the false positive rate for the images that do not present the skin at all.

References

1. Bilal, S., Akmeliawati, R., Salami, M.J.E., Shafie, A.A.: Dynamic approach for real-time skin detection. J. Real-Time Image Process. (2012)
2. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: A skin tone detection algorithm for an adaptive approach to steganography. Signal Process. 89(12), 2465–2478 (2009)
3. Chen, Y.H., Hu, K.T., Ruan, S.J.: Statistical skin color detection method without color transformation for real-time surveillance systems. Engineering Applications of Artificial Intelligence 25(7), 1331–1337 (2012)
4. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proc. Int. Conf. on Machine Learning, pp. 233–240 (2006)
5. Greenspan, H., Goldberger, J., Eshet, I.: Mixture model for face-color modeling and segmentation. Pattern Recognit. Lett. 22, 1525–1536 (2001)
6. Jiang, Z., Yao, M., Jiang, W.: Skin detection using color, texture and space information. In: Proc. IEEE Fuzzy Systems and Knowledge Discovery, pp. 366–370 (2007)
7. Jones, M., Rehg, J.: Statistical color models with application to skin detection. International J. of Computer Vis. 46, 81–96 (2002)
8. Kakumanu, P., Makrogiannis, S., Bourbakis, N.G.: A survey of skin-color modeling and detection methods. Pattern Recognit. 40(3), 1106–1122 (2007)
9. Kawulok, M.: Fast propagation-based skin regions segmentation in color images. In: Proc. IEEE Int. Conf. Automat. Face and Gesture Recognit., pp. 1–7 (2013)
10. Kawulok, M., Kawulok, J., Nalepa, J.: Spatial-based skin detection using discriminative skin-presence features. Pattern Recognit. Lett. 41, 3–13 (2014)
11. Kawulok, M., Kawulok, J., Nalepa, J., Papiez, M.: Skin detection using spatial analysis with adaptive seed. In: Proc. IEEE Int. Conf. Image Process., pp. 3720–3724 (2013)
12. Khan, R., Hanbury, A., Sablatnig, R., Stoettinger, J., Khan, F., Khan, F.: Systematic skin segmentation: merging spatial and non-spatial data. Multimedia Tools and Applications 69(3), 717–741 (2014)
13. Kovac, J., Peer, P., Solina, F.: Human skin color clustering for face detection. In: Computer as a Tool, vol. 2, pp. 144–148 (2003)
14. Lee, J.S., Kuo, Y.M., Chung, P.C., Chen, E.L.: Naked image detection based on adaptive and extensible skin color model. Pattern Recognit. 40, 2261–2270 (2007)
15. Phung, S.L., Chai, D., Bouzerdoum, A.: Adaptive skin segmentation in color images. In: Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process., pp. 353–356 (2003)
16. del Solar, J.R., Verschae, R.: Skin detection using neighborhood information. In: Proc. IEEE Int. Conf. Automat. Face and Gesture Recognit., pp. 463–468 (2004)
17. Wang, X., Zhang, X., Yao, J.: Skin color detection under complex background. In: Proc. IEEE Int. Conf. Mechatronic Science, Electric Engineering and Computer, pp. 1985–1988 (2011)
18. Yang, G., Li, H., Zhang, L., Cao, Y.: Research on a skin color detection algorithm based on self-adaptive skin color model. In: Proc. IEEE Int. Conf. Communicat. and Intellig. Informat. Security, pp. 266–270 (2010)
19. Yogarajah, P., Condell, J., Curran, K., Cheddad, A., McKevitt, P.: A dynamic threshold approach for skin segmentation in color images. In: Proc. IEEE Int. Conf. Image Process., pp. 2225–2228 (2010)
20. Zhu, Q., Cheng, K.T., Wu, C.T., Wu, Y.L.: Adaptive learning of an accurate skin-color model. In: Proc. IEEE Int. Conf. Automat. Face and Gesture Recognit., pp. 37–42 (2004)